COUNTING WHAT COUNTS:
USING BIG DATA TO DRIVE SOCIAL IMPACT

Sonali Arora

TC 660H
Plan II Honors Program
The University of Texas at Austin

May 8, 2017

Angela Newell
McCombs School of Business
Supervising Professor

James Howison
School of Information
Second Reader

Abstract

Author: Sonali Arora

Title: Counting What Counts: Using Big Data to Drive Social Impact

Supervising Professors: Angela Newell, James Howison

As data accumulates and the world becomes increasingly interconnected, the resulting "big data" that is generated offers potential for applying this information towards social change. The private sector has been successful in utilizing big data analytics for driving innovation and economic growth. Gradually, the social sector is adopting these practices and taking advantage of the vast amount of data available to tackle complex societal problems and assist underserved populations. By incorporating data-driven analysis into efforts aimed at driving social impact, organizations across sectors can facilitate the development of innovative solutions to the problems society needs answers to. Despite this potential, numerous challenges stand in the way that have prevented the social sector from making all these possibilities a reality, as this type of work requires a collaborative model given the dynamic nature of social problems and the complexity of their solutions. What is needed in order to accelerate the use of big data solutions towards social innovation? This thesis explores the current state of big data in the social sector, examining case studies of successful organizations that have used big data in addressing the most pressing issues in society, and the impediments that have prevented similar data-driven solutions from becoming more widespread. This thesis aims to provide direction in guiding efforts going forward to expand the use of big data in driving social impact, offering suggested strategies to develop sustainable solutions in light of the obstacles impeding such efforts.

Acknowledgements

I would like to thank my supervising professor, Dr. Angela Newell, and my second reader, Dr. James Howison, for their dedication and support throughout this process. This thesis would not have been possible without their valuable input and guidance. I would also like to give a sincere thank you to my parents for their love and support, and to my family and friends for their constant positivity and words of encouragement.

Table of Contents

Introduction

With the massive volume of data in today's world and the technological advances that accompany it, the rise of big data offers potential for data-driven solutions to generate positive impact. As the amount of data accumulates at a rapid speed, how can nonprofit organizations and social enterprises benefit from big data? Driven by developments in technology, the global spread of devices has propelled the dissemination of information as the world becomes continuously more digitized. As people, their devices, and objects grow increasingly connected, more data is collected at faster rates, allowing information to be analyzed on a larger scale than ever before (World Economics Forum, 2015, p. 6). The immense volume of data is only growing – it is estimated that "global data traffic will cross 100 zettabytes per year by 2025" (Frost & Sullivan, 2015, p. 3). Combined with technical advances that make it possible to analyze, this deluge of data, known as "big data," holds promise for social impact through informed decision-making. Proponents of big data claim that by making smarter data-driven decisions, companies, individuals, and governments can understand people's preferences, behaviors, health, and more (World Economics Forum, 2015, p. 6). From improving healthcare to preventing crime to measuring poverty, the applications of big data in today's world are widespread ("Data, data," 2010).

One of the most promising applications of big data is in the social sector, where nonprofit organizations, government agencies, and other socially conscious enterprises collaborate to address the needs of society and develop solutions for underserved populations. When it comes to identifying major challenges and knowing where to focus efforts, big data analysis offers a powerful tool for informing decisions (Bloomberg Philanthropies, 2016). However, a number of challenges stand in the way of realizing the potential for big data in improving society. While

some efforts have been successful in utilizing big data to develop innovative approaches for understanding social issues, the power of big data remains largely untapped (World Economics Forum, 2015). Through case studies of successful big data efforts in different areas, this paper aims to illustrate the potential for big data in solving the most pressing issues in society, and thus the need for more coordinated strategies for improvement.

The metrics used for measuring the success of big data vary depending on the type of social issue the data supports. As an example, in addressing public health issues, success for disease outbreak containment could be measured by the amount of time a disease is contained, while efforts to improve access to health care would likely be evaluated based on the percentage of the population that has access to health clinics and facilities (Bloomberg Philanthropies, 2016). As the case studies will illustrate, finding metrics to quantitatively measure the impact of big data initiatives on society proves challenging for many organizations. Efforts to use big data in analyzing social problems are individual to a specific issue or region, and therefore the metrics used to evaluate their effectiveness are unique to the situation.

In order to address the issues of advancing the use of big data, this thesis focuses on the current and potential applications of big data for social innovation, the barriers to overcome, and the issues associated with developing data-driven solutions to social problems. One major issue identified in the study is the need for increased collaboration in implementing these strategies, as innovation requires improved access to data. Much of the data that could be used for analysis is already being generated, so it is primarily a matter of making better use of it through increased access (World Economics Forum, 2015, p. 4). As this thesis will demonstrate, sharing data is a significant step towards maximizing its potential given the cooperative nature of applying big data solutions in the social sector. These issues will be explored in further depth and illustrated

through a comparative case study of organizations that have been successful in their use of big data.

This thesis will begin with a background section that defines big data, illustrates its potential in the social sector, and discusses the primary challenges currently impeding many organizations. The methodology section provides an explanation of the case study approach and goals of the design, followed by a case studies section that explores the big data initiatives of four organizations in detail. The analysis section summarizes the findings of these case studies and offers recommendations going forward for accelerating the use of big data in the social sector in light of the major challenges identified. Finally, this thesis concludes with a summary and discussion of this research, the implications of these findings, and the future of big data in addressing the needs of society.

Background

What is Big Data?

The concept of big data refers not only to the rapid influx of data, but also the advances in data analytics that allow us to make sense out of complex, massive data sets. Big data is frequently defined by the "Four Vs": volume, variety, velocity, and veracity (World Economics Forum, 2015, p. 6). Immense volumes of data are collected as the amount of information grows at a rapid rate, demanding increased data storage (Roski, Bo-Linn, & Andrews, 2014). Data exists in a variety of forms – from images to GPS signals to text messages – and comes from a range of sources. Each individual leaves a trail of data through numerous personal devices, and companies collect data in various ways, including through social media, customer transactions, and online advertising. Along with the volume of data, the speed, or velocity, of creating and

processing data is equally advanced. The potential for real-time analysis allows for flexibility in managing data and speed in generating results (McAfee & Brynjolfsson, 2012). Finally, veracity applies to the uncertainty of data - data is often messy, but analytics technology for big data can make use of this data, despite the noise (World Economics Forum, 2015, p. 6).

Data-driven solutions have proved successful in the business world. Realizing the potential this influx of data offers for driving innovation, firms have been quick to adopt big data into their business practices ("Data, data," 2010). With this comes a demand for data management tools to make sense of data. Not surprisingly, the analytics software industry is "estimated to be worth more than $100 billion and growing at almost 10% a year, roughly twice as fast as the software business as a whole" ("Data, data," 2010). While not all efforts are met with success, many companies that make use of big data effectively have a significant competitive advantage. Improving efficiency, reducing costs, and targeting consumers are a few examples of the benefits. One study found that companies that were the most advanced in using big data in their industries were on average "5% more productive and 6% more profitable than their competitors" (McAfee & Brynjolfsson, 2012, p. 63-64).

For online retailers like Amazon, statistics and data mining tools have transformed the way companies understand consumer preferences and shopping behavior. By analyzing clicks, past purchases, transactions, and more, retailers can predict the products customers are likely to purchase in advance, and then use these insights in developing targeted marketing strategies. Firms across many industries are making use of similar strategies, generating predictions from data through algorithms to make decisions in all areas of the business (McAfee & Brynjolfsson, 2012, p. 62). For example, credit card companies use data to generate predictions in order to determine default risk of loans (Lippert, 2014). Google is developing algorithms using search

data to understand translation and speech patterns (Markoff, 2012). The applications are widespread across industries, creating value for firms through diverse approaches.

In addition to businesses, big data plays a significant role in health care as well. The health care industry heavily invests in information technology (IT), so there is more potential than ever for big data in facilitating improvements in the health of populations. The shift towards increasingly digitized health records and research databases has resulted in advances in analytics that make it possible to increase efficiency and reduce cost while improving the quality of care. In fact, Cleveland Clinic listed big data as one of the top recent medical innovations (Cleveland Clinic, 2011). One of the most significant breakthroughs has been the use of data mining in personalized medicine. Automated evaluation of MRIs, CT scans, and x-rays from a patient's individual profile can be used in diagnosis and treatment by developing automatically generated personalized care for patients (Roski et al., 2014).

When it comes to the private sector and developing innovative solutions, organizations in many industries are approaching a point where they will not be able to survive without big data (Frost & Sullivan, 2015). Growth in big data in the social sector, however, has progressed more slowly. To understand how big data can be harnessed towards improving lives, the next section will provide background on how big data can address social issues.

The Social Sector

Before considering the applications of big data for social innovation, it is important to clarify what exactly the social sector refers to and the types of entities it encompasses, as well as trends that have set the scene for the use of big data. The social sector is comprised of individuals and organizations that provide goods and services for social and environmental issues

that are not addressed through the mainstream economy. Often referred to as the third sector because it differs from the public sector and private sector, this sector includes nonprofits, philanthropic organizations, civic organizations, socially conscious businesses, governments and government agencies, and collaboration across these entities. The U.S. social sector is growing rapidly and is made up of over one million organizations that tackle issues ranging from health and education to poverty and public safety. The social sector makes up roughly six percent of the country's GDP and continues to grow as new products and services are offered (McKinsey, 2014). Additionally, private giving from businesses, foundations, and individuals has grown substantially – total private giving in 2014 amounted to $358 billion in the U.S. and continues to increase each year. Along with the rise in donations is a growing interest in civic engagement and volunteering among citizens. In 2014, approximately 25% of the adult population volunteered or gave time to an organization, totaling nearly 9 billion hours (Urban Institute, 2015).

A significant trend recently has been the role that businesses and for-profit entities have started to play in social innovation and community impact, as this type of work is no longer restricted to the nonprofit world. Many businesses have started to place a higher priority on social impact, developing new corporate social responsibility programs and initiatives. B Corporations (certified for-profit companies committed to creating social and environmental value), social enterprises, and socially conscious for-profit companies make up a portion of the social sector that is rapidly growing (McKinsey, 2014). One result of the intervention of private sector businesses is the increasing application of technology towards social impact. As large businesses take an increasing interest in social responsibility, tackling issues that have traditionally been addressed primarily by nonprofits, many of these companies have started to

apply strategies and resources from their for-profit operations into these social impact initiatives (Frost & Sullivan, 2015).

Along with these technological advances is a growing pool of skilled experts. The interest in these areas, both in technology and in the desire to create social innovation, has resulted in the development of new university programs, courses, and training. The result of this growing interest has been that more people are being trained in the skills necessary to do this type of work, applying technical skills towards social problems. Data analysis for social innovation requires skills beyond just an understanding of technology, coding, and numbers. Many organizations have emphasized the importance bringing together people from different disciplines and sectors to create synergy and tackle problems from different perspectives. This growing pool of individuals trained in data science and technology with an interest in applying these skills towards solving social problems has created a group that fits this profile (Data on Purpose Panel, 2016). While there are more factors at play, these themes have allowed for the promotion of big data towards solving social problems, which have not traditionally been dealt with through significant use of technology. This trend has created an environment in which big data or any type of data-driven solutions can be produced, given both the access to technology and the growing talent pool. This rise in digital and systems-based approaches creates an opportunity for large-scale big data initiatives in the social sector.

Big Data for Social Innovation

The big data revolution exists beyond just the private sector, and well beyond developed countries. Big data has the potential to improve the lives of people globally and drive social innovation. Defined as “the deployment of technology and new business models to bring about

real, positive change to the lives of individuals and societies, creating shared value," social innovation is the process of developing novel solutions to the most complex societal problems (Frost & Sullivan, 2015, p. 3). Big data analytics can lead to improved understanding of a population's behavior, needs, and resources, offering a powerful tool for implementing policies and programs aimed at bettering society (UN Global Pulse, 2013).

Both the science community and the business world have been fast to adopt big data practices and make use of data-driven decision-making, but nonprofits, governments, and organizations committed to social impact have lagged in comparison. Recent trends in technology, particularly in developing countries, have resulted in a substantial increase in data produced. For example, the spread of mobile devices across the developing world is a significant development, as phones can produce and track data on billions of people who previously had no access to such technology (UN Global Pulse, 2013, p. 2). The data collected from devices can be a powerful tool in implementing programs aimed at bettering society. Despite this potential, the social sector has yet to make all these possibilities a reality, as numerous challenges exist that stand in the way of harnessing big data for social innovation. Developing technology to collect data, facilitating the exchange of data across sources, and ensuring proper use of data are all difficulties that must be addressed in order to enable success (United Nations, 2014).

Despite the barriers, many initiatives involving big data in the developing word have been successful, holding considerable promise for future efforts. For example, social media analytics allowed for the development of a food price index in Indonesia. Looking at Twitter data, a collaborative project between United Nations Global Pulse, the World Food Programme, and the Indonesian Ministry of National Development Planning discovered that tweets about food prices accurately reflect real prices. From this, they created a technology that takes food

prices stated in tweets and produces a food index in real-time. This same method for data mining can be applied to any pricing or location, and make use of other social media platforms. Another successful initiative is an educational program in Rio de Janeiro, in which cameras were attached to kites to collect images of the cities infrastructure, with the aim of teaching youth how to prepare for natural disasters and understand the risks. This led to efforts to repair various facilities and remove barriers that could complicate evacuation (United Nations, 2014, p. 13).

Another area in which big data has proved successful is public health. As health care industries across the world grow increasingly digitized, big data has promising applications in public health, especially in developing countries where addressing disease outbreaks is a major focus of public health initiatives. One example is the Mtrac program in Uganda, aimed at treating and preventing malaria. As one of the leading causes of death, malaria takes up a substantial amount of health care resources in developing countries, which devote a large amount of effort towards preventing and containing outbreaks. Mtrac makes use of mobile phones by allowing health workers to submit surveys via SMS to notify officials of outbreaks. The program creates a means for communicating necessary information, such as medicine availability or resource shortages. The program has brought dramatic improvements and has grown extensively, as thousands of workers communicate and data is captured at a low cost without delay. Through this communication, the amount of health facilities without enough Artemisinin-based Combination Therapies (ACTs), used for malaria treatment, has been reduced by 65%, down to 15% from 80% originally. This program demonstrates how communication-based data and analytics tools can dramatically improve the effectiveness of treating disease outbreaks (United Nations, 2014, p. 9).

While these solutions sound promising, there is still substantial room for growth. Significant challenges stand in the way of increasing the use of big data for social impact, but are being met with efforts to overcome barriers to improvement. The next section highlights the key issues involved with big data that interfere with the implementation of more efforts involving data analytics.

Challenges

A number of challenges stand in the way of utilizing big data for social impact. When it comes to generating data-driven solutions for social issues, complexities in data greatly limit the number of initiatives taken and the effectiveness of such efforts. In low- and middle-income countries, these challenges are further complicated by technological barriers, which increase the difficulty of gaining access to data (United Nations, 2014, p. 9). Different layers of challenges exist, as there are multiple stages in which complications arise – data collection, data access, and data use (World Economics Forum, 2015, p. 8). Challenges at each stage are briefly examined below.

Collection

The first challenge involves data collection and developing technology to store and collect. Given the increasing digitization of business and health care, more and more data is generated, as information is electronically stored. However, difficulties in collecting data, such as data being stored in the wrong format or getting lost in administrative systems, have stifled this growth. Despite technological developments, many countries still lack the basic infrastructure needed for creating and testing plans. Proper software may be lacking, or

incompatible for combining sets of data (Panhuis et al., 2014, p. 3-4). In some cases, the cost of adopting new technology is too high, or lack of Internet access complicates data collection and analysis (United Nations, 2014, p. 6-8).

Access

The most significant impediment involves data access and developing strategies for data sharing. More and more data is generated, but only a fraction of this data is made available for analysis. This gap between data generated and data made accessible impedes efforts to develop data-driven initiatives. Most opportunities for harnessing big data in improving social and health issues involves data from a variety of entities across multiple sectors and countries, given the collaborative nature of these solutions. These different players, however, frequently lack the means or the incentive to share their data (United Nations, 2014, p. 15).

A major issue is that the public and social sectors lack the same access that the private sector has to data. Companies collect and accumulate massive amounts of data, both for their own use and "as a byproduct of other transactions" (World Economic Forum, 2012, p. 3). Though they may not realize it, much of this data has the potential to benefit the public. Private sector firms own information that is valuable for solving social problems, but rarely share this data. As a result, the power of big data remains untapped because data that could be analyzed is kept privately owned (Stempeck, 2014).

Companies are not willing to share their data for a variety reasons. One reason is due to a lack of awareness, as they may not realize the potential their data can have in bringing about social change (Stempeck, 2014). Additionally, they may worry about transparency and the implications of making data openly available when it comes to factors such as privacy and

competition. Ultimately, data is a resource and companies see their data as proprietary, believing their data offers them a competitive advantage. Another concern is privacy, as firms may hesitate to share data on the basis that doing so threatens the privacy of their consumers, from whom the data was collected. The World Economics Forum, in 2012, laid out these issues, claiming that they are legitimate worries and without taking proper steps to mitigate these concerns, this barrier to data sharing cannot be overcome.

These concerns are not just held by companies but by other institutions as well. Nonprofits and social impact organizations have similar worries, though generally on a smaller scale, and research organizations can be equally unwilling to share data. Another major impediment to sharing is that the dynamic of business and health care organizations does not provide for an easy process to share data. Data is often siloed, largely because organizations have not developed processes for combining data and communicating outcomes. Generally, the potential for big data solutions cannot be realized without a culture that revolves around information management (Groves, Kayyali, Knott, Van Kuiken, 2013).

Strategies for coordinating the exchange of data are needed to overcome this obstacle of data access, as transferring and compiling data requires a shift from a culture of withholding data, in business and in health institutions, to a culture of opening up access to data. Moving forward, steps will need to be taken towards developing standards for regulating data and governing the use of shared data (Brack & Castillo, 2015).

Use

Finally, issues arise from the concern that data will not be used in the way it is intended to. This can promote a lack of trust between data owners, as one party may not be willing to give

out their data due to uncertainty in how this data will be used. For example, a mobile provider sharing call data with a company may be concerned with how a company will anonymize the data and protect its customers' personal information. Another example is health data, where there is a large concern over safety issues due to the possibility of potential abuse of medical information. There is always the risk that big data can be exploited for financial gain or towards initiatives that are unfavorable or wasteful (Panhuis et al., 2014, p. 5).

Combined, these obstacles stand in the way of making use of potential big data solutions for social innovation, reducing the efficiency and capability of big data in decision making. What policies and standards can be established to accelerate the use of big data? What are the best solutions for encouraging cross-sector collaboration? The next sections explore these questions and seek to develop strategies for improvement by drawing from successful case studies.

Nonprofits and Technology

Before evaluating the social sector's ability to adopt more data-driven strategies, it is important to consider the nature of nonprofit funding and the resulting limitations placed on overhead spending. As mentioned previously, one major challenge for nonprofits is the lack of infrastructure, as many do not have the capacity to make significant investments in technology. It is common for funders to offer financial support for direct services, providing funding for direct expenses relating to specific programs or projects. The result of this restricted funding is that capacity building and indirect expenses frequently go unfunded (Gregory & Howard, 2009).

Restricted funding is common in the nonprofit sector, as funders often provide grants directed at specific programs, placing limits on the type of costs these funds can be used towards. This funding practice is fueled by the misleading belief that nonprofits that spend less on

overhead costs must spend more on programs. Many have argued that this type of funding severely limits the impact of nonprofits, making them inflexible once they accept grants with restricting conditions (Shoemaker, 2015).

A harmful cycle exists, known as the "nonprofit starvation cycle," that frequently prevents nonprofit organizations from investing in the infrastructure they need in order to function effectively (Gregory & Howard, 2009). Funders have unrealistic expectations when it comes to the costs nonprofits incur. Donors expect to see high performance and low overhead spending, without actually funding overhead expenses. This leads to nonprofit organizations spending less than they should on overhead and underreporting their spending, which further fuels the expectations of funders and leads to a vicious cycle of nonprofits failing to receive proper funding for fixed costs and capacity building (Gregory & Howard, 2009).

As a result, nonprofit organizations frequently do not spend on resources they would benefit from. This leads to shortages in spending in a variety of areas, including information technology, staff training, and salaries. For example, many nonprofits face issues in hiring because they are unable pay specialists a competitive salary. Additionally, the technology and infrastructure these organizations could be investing in would aid them in tracking progress and measuring performance. This furthers the issue, since funders want to see performance of programs before they will fund general spending like overhead, but this overhead spending is needed in order to both achieve and measure that performance (Gregory & Howard, 2009).

Methodology

In order to understand the challenges that stand in the way of nonprofits, companies, and organizations devoted to social impact from using big data, a comparative case study was conducted. Four organizations that have been successful in their use of big data were analyzed to understand the challenges faced by these organizations and to propose strategies to overcome such obstacles. A comparative case study was selected as the method for analysis due to the diverse nature of big data initiatives in the social sector. This paper seeks to reveal the major challenges in this area and to offer future guidance by analyzing the similarities and differences across four different organizations, identifying patterns that can be applied going forward. The factors for comparison are primarily qualitative, given the dynamic nature of big data and social sector trends. Because the complexity of the issue requires exploratory research, in terms of understanding the both the challenges and potential for big data for social impact, a qualitative approach was taken in analyzing the case study findings. Given the lack of homogeneity in this area, as organizations and initiatives using big data for social innovation vary significantly in terms of type, size, and issue addressed, a case study analysis sheds light on how these differing contexts impact big data initiatives. The variety in the organizations chosen provides a representation of the heterogeneity that exists in reality. Despite this variety, however, there are shared characteristics that exist across organizations and big data initiatives, and this study aims to highlight the commonalities and differences in order to draw lessons from both.

In order to select a representative pool of organizations, the four case studies chosen vary both in terms of size and structure, as well as type of issue addressed. To determine which organizations to include as case studies, organizations were first grouped based on characteristics: type of initiative, type of company or organization (nonprofit vs for-profit),

geographic region served, and what type of partners the organization collaborates with. From these groups, four final groups were created, and one in each group was interviewed: 1) large companies, 2) nonprofit organizations, 3) hospitals with analytics-driven programs, and 4) universities that fund big data projects. The interviewed organizations are outlined below:

For-Profit Company	**Hospital**	**Nonprofit Organization**	**University**
IBM	Cook Children's Center for Prevention of Child Maltreatment	Children's Optimal Health	Austin Area Sustainability Indicators Project

The four organizations interviewed were IBM, Children's Optimal Health, Cook Children's Center for Prevention of Child Maltreatment, and Austin Area Sustainability Indicators Project. Interviews were conducted with representatives from each of these organizations to gain insight on their big data initiatives. The interviews included one person from each organization who played a role in the big data projects, for a total of four interviews. The following questions were discussed with each interviewee:

- *Where does the organization get its data?*
- *What are the major challenges?*
- *How is the data kept private and secure?*
- *What metrics are used to measure the success of initiatives?*
- *What are the organizations next steps in using data?*

Each of the case studies will be explored in detail in the following section, comparing the factors listed above across each of the organizations. From these case studies, this paper will propose a

set of recommendations that address the challenges involved with using big data for social innovation.

Case Study 1: Children's Optimal Health

Organization Overview

Children's Optimal Health is a nonprofit organization in Austin, Texas dedicated to improving the health of central Texas youth through data visualization. By providing maps to visualize neighborhood data and assess health needs, Children's Optimal Health aggregates data from a variety of sources to inform decision makers on how to effectively target resources. Using Geographic Information System (GIS) mapping, Children's Optimal Health provides accurate and frequent data to promote new policies, research, and community engagement (Children's Optimal Health, 2017).

The maps Children's Optimal Health provides are for the Austin community to improve the understanding of children's health. Austin is one of the fastest growing cities in the country, and as the city continues to grow and housing prices rise, many low-income families face financial struggles. The goal of Children's Optimal Health is to identify issues and solutions to major health concerns, and to equip these families with the right information through data-driven analysis.

Children's Optimal Health started as a collaborative of institutions in 2006 devoted to advancing the health of Austin youth through collective resources, coordination, and community engagement. The main focus areas included organizing the use of current resources, educating the community, finding innovative ways to use technology, and researching the best solutions for health issues in the community. In 2008, Children's Optimal Health grew from this collaborative

network, as five board members gained funding, founded the nonprofit, and committed themselves to finding innovative ways to promote solutions to the most significant health issues in Austin.

Data Analysis

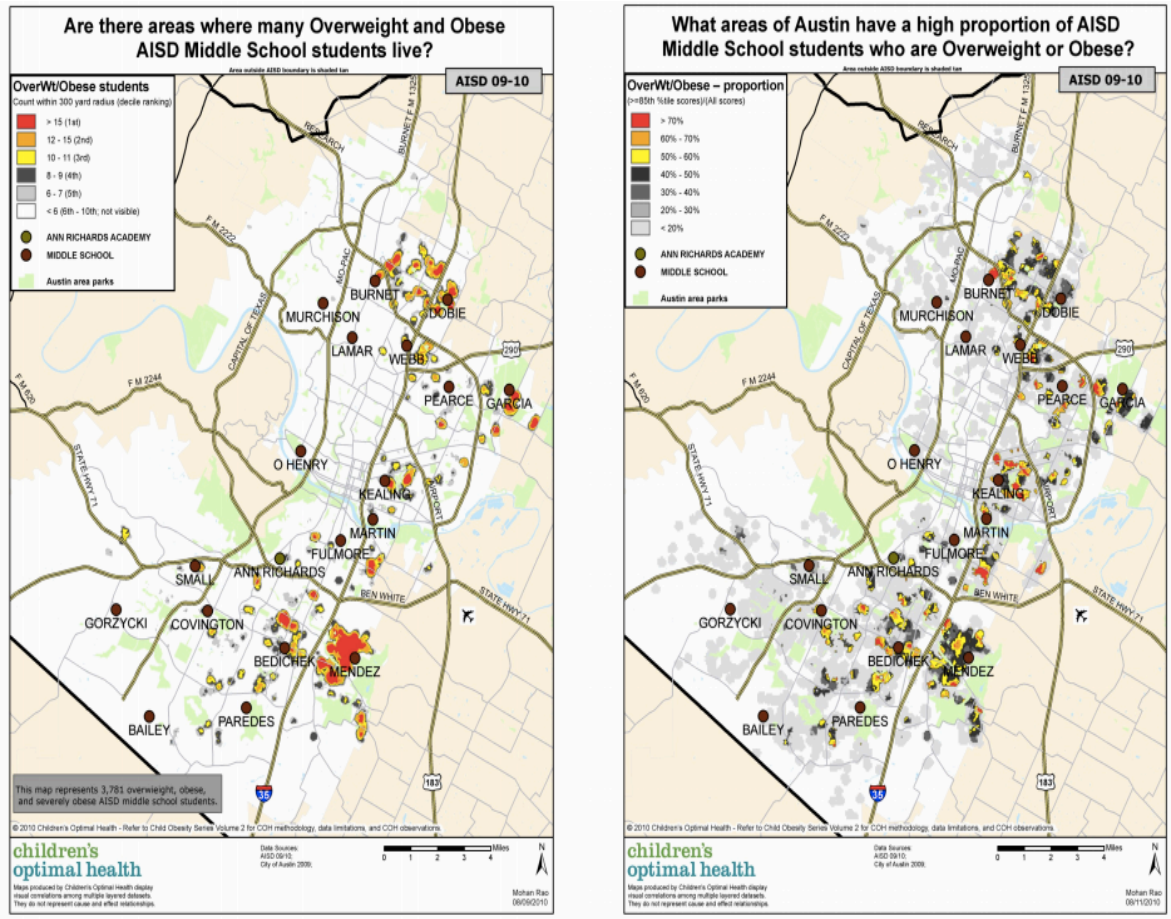

Figure 1. Child obesity among AISD middle schools. The graphs above illustrate the areas with the highest rates of obesity among AISD students (Children's Optimal Health, Child Obesity, 2011).

Children's Optimal Health focuses on data visualization as a tool for assisting the community in understanding the health needs of Austin's youth. Through Geographic Information System (GIS) mapping, Children's Optimal Health develops maps that they claim

offers a powerful tool by conveying neighborhood data in a manner that is easy to visualize. This spatial analysis method provides an illustration of shifting community needs and health services available. These maps take into account a variety of demographic data and health data to provide insight on where the hotspots are in terms of numerous health-related issues, illustrating where community partners should target resources. Once hotspots are identified, the maps can be analyzed further to look at specific factors, combining data on all aspects of the community. These factors include everything from schools, transportation, and healthcare availability, to socioeconomic status and race. By considering health in light of these factors, these maps can be used to understand dynamic trends in the community and identify patterns that may have gone unnoticed. The maps provide a data-driven representation of youth health issues and can be easily understood despite the volume and complexity of data required to develop them.

Data Sources

Children's Optimal Health reported developing Data Sharing Agreements (DSAs) with organizations in Central Texas to collect data needed to develop maps. The first step in developing these agreements is to get a business agreement from the organization and then data use agreements that Children's Optimal Health negotiates depending on the project. For example, if Children's Optimal Health is working with Seton Healthcare Family to acquire data on children's asthma, they will begin by establishing business and data use agreements, which allows them to clarify how the data is going to be used. After this, they can get the data pulled based on what data they need. Children's Optimal Health also works with insurance companies, which provide information on diagnostic codes, medications, and whether or not prescriptions

have been refilled. This has been used for getting information on inhaler use, since asthma is one of the major health concerns among children.

Children's Optimal Health receives yearly datasets from local school districts including Austin Independent School District. This data includes fitness testing information, such as body mass index (BMI) and cardio fitness metrics. The data also comes from Public Educations Information Management System (PEIMS) data, which includes information on students' demographics, academic performance, and financial measures such as free or reduced meals. Combined, this information provides insight on factors like poverty and attendance rates, which can be applied to understanding and improving health issues.

In addition to school districts, Children's Optimal Health also makes use of public datasets. Austin Police Department was cited as a significant source of data. In acquiring police data, arrests and other irrelevant information can be factored out. Police data can be used in a number of ways to understand children's health. An example is transportation child injury, as police data can provide information on where accidents occur and citations for kids not restrained properly.

In acquiring data from these sources, Children's Optimal Health argues that it is successful due to its ability to develop trust with partner organizations that provide data. Through these close relationships, Children's Optimal Health has established data sharing agreements that form the basis of the work it does, allowing the organization to gain the specific data it needs to provide a full picture of health risks in the Austin area.

Challenges

Children's Optimal Health identified acquiring data and keeping it secure as one of the most significant challenges it faces. There are a variety of difficulties involved in negotiating data sharing agreements, as it is essential to ensure security throughout the process and make other organizations feel comfortable sharing their data. The organization believes its biggest strength has been its ability to develop trust and convince partner organizations to share their data. It began with a few individuals at Austin Independent School Districts who were willing to take the leap, and since then has grown into a network of organizations providing data that feeds into the neighborhood maps Children's Optimal Health develops.

In addition to data sourcing, measuring the success of its maps is another challenge cited by Children's Optimal Health that the organization faces. Although the maps are clearly beneficial to the community, it can be difficult to specifically measure the impact of the maps using precise metrics. Specifically, it is challenging to isolate and measure the direct impact of the maps given that the maps are part of a community wide effort to address issues relating to children's health. However, Children's Optimal Health is developing new strategies to tackle this challenge and translate its success into metrics.

Security

The data Children's Optimal Health collects and transforms into maps is protected throughout the process, including all personal information associated with the data. When data is in transit, it is encrypted to ensure security. Currently, Children's Optimal Health is in the process of moving all its secure data to the Texas Advanced Computer Center, a research center with advanced computing technology that will provide maximum security. Depending on where

the data is sourced from, it may be de-identified already. For example, AISD has a specific identification number for Children's Optimal Health so that student data can be tracked over time without knowing names.

Alternatively, the data it receives may be reflected through points on a map. For rooftop level data, the organization has developed algorithms that slightly shift the data points so they stay within a neighborhood but are somewhat jittered. These data points are then combined and transformed into a map that provides visualization of rooftop level data by neighborhood, with no way of finding the identity of individual people. Providing maps based on personal-level data rather than a higher level of data allows Children's Optimal Health to find hotpots it otherwise would not have been able to. County level data, for example, would not provide a sharp enough point where it is possible to fully understand what is occurring in a particular area. By using rooftop level data to analyze factors such as public transportation transit lines, crime rates, or poverty rates in different areas, the data points do not diffuse over a large geographical area and allow for the discovery of hotspots that can be identified by neighborhood.

Success Metrics

According to Children's Optimal Health, measuring the impact of data initiatives poses a challenge. The maps the organization develops are released to the community, so anyone can go on their website and access the maps. Using these maps, health organizations have used the maps to understand where to focus resources, new school programs have been established to address health needs, and new clinics have been opened based on hotspots. As a result, there is no question that these maps have proved beneficial to the community. However, this success can be difficult to translate into metrics.

Children's Optimal Health hopes to partner in more research-based projects in the future. As that happens, the organization will have metrics that come from that program piece. However, the organization requires outcomes from different sources using the maps. There is no way of measuring the effect since no two individual programs count the same factor in the same way. Up to this point, the organization believes that using a time series has proven to be the best way for them to measure impact. The data it receives from Austin Independent School District, including student's BMI and cardio data, have seen changes over time. Additionally, new school programs such as Hot Sports, targeted at improving children's health through physical activity, have been implemented and have led to improvements.

Next Steps

Going forward, Children's Optimal Health is very focused on population health projects. This work will involve analyzing health, education, and social determinants in addressing population health issues. Many problems cannot be solved by only addressing a single factor, but rather require looking at a variety of areas and collaborating across service providers. For example, in the context of medicine, there are many situations where doctors can provide patients with medication to avoid symptoms without actually treating the root cause due to missing information. Population health projects and research aim to close this gap by encourage groups to work collaboratively and bringing different areas of study together to create a more comprehensive understanding of health.

Currently, the organization is in communication with a medical school and its population health department. Children's Optimal Health has also been a part of a research project with Dell Children's Medical Center. According to the organization, partnerships will take time to develop

but will prove beneficial to both Children's Optimal Health and the organizations it partners with. They claim that because the organization does not provide direct services, partnerships are all the more important. For nonprofits that offer indirect services, like Children's Optimal Health and its neighborhood maps, obtaining funding can be difficult since there are no exact metrics to provide funders. Getting the community more interested in data is another focus of the organization going forward and is it continues to expand its projects and outreach.

Case Study 2: Cook Children's Center

Organization Overview

Tarrant County, located in the city of Fort Worth, has one of the highest child abuse rates in Texas. Compared to the city of Houston, which has a similar population of children, the Dallas-Fort Worth area has nearly twice as many child abuse incidents. Doctors at Cook Children's Hospital in Fort Worth are actively working to solve this issue and bring down these rates. The Center for Prevention of Child Maltreatment is a center funded by Cook Children's Health Care System with this goal in mind (Tarrant County, 2015).

It remains unclear why the Dallas-Fort Worth area experiences such high rates of child abuse. Because of the high occurrences, Cook Children's trains doctors and health care professionals to recognize the early signs of abuse. However, detecting child abuse can be challenging, especially when children are at a young age and may not interact with adults other than their parents. Because of this challenge, Cook Children's began focusing on prevention and started utilizing big data technology to predict the neighborhoods where cases of child abuse are most likely to occur. Predictive analytics has been used in addressing a variety of issues, such as detecting crime areas or predicting where shootings will occur, but it had not been applied

towards domestic violence before. By applying this technique, Cook Children's has been able to predict with high accuracy the neighborhoods where child maltreatment will occur for a given year, and use that information to allocate resources to prevent occurrences in these areas (Using Data to Predict, 2016).

Data Analysis

Cook Children's initially collects data on a variety of risk factors such as poverty levels, parental substance abuse, and domestic violence. Cook Children's then uses free technology developed by Rutgers University, known as Risk Terrain Modeling (RTM), to analyze the numbers and develop a forecast for locations where child abuse is most likely to occur. The locations identified are specific areas within neighborhoods, providing detailed insight on where to target prevention efforts. Understanding where the high-risk areas are creates an opportunity to provide prevention services to the areas that require them most.

Risk Terrain Modeling was initially developed in the context of public security, and used for applications in criminal justice. Given the success it had in predicting robberies, shootings, and other crimes, the technology became widely used in criminology. This is the first application of Risk Terrain Modeling towards child maltreatment, and the predictive ability of the model provides a tool for directing prevention efforts. What makes Risk Terrain Modeling different from other methods of hotspot mapping is that it develops predictions of future occurrence locations based on the cumulative collection of risk factors, rather than reflective analysis of past instances. As Cook Children's describes, this type of geospatial analysis offers a number of advantages, as it considers the relative spatial effect of a variety of factors on the dependent variable, in this case child abuse, increasing the predictive power of the model. The identification

of these areas offers a data-driven recommendation to guide prevention efforts (Child Abuse & Neglect, 2016).

In terms of funding, the Center is funded by Cook Children's Health Care System. This funding goes toward personnel, which is a large portion of total costs, and software and program licenses. As mentioned, the Risk Terrain Modeling software is provided free of charge, and Cook Children's has very low data costs. Given the close relationships it has established with organizations that provide data, Cook Children's receives data for free, aside from very nominal processing fees.

Data Sources

In order to assess the factors associated with child maltreatment, Cook Children's gathers data from a variety of sources to properly examine these factors. A major source of data for Cook Children's is the Fort Worth Police Department (FWPD), which keeps publicly available data but requires a request submitted to the police department. Through this data, Cook Children's can understand where aggravated assault is likely to occur in terms of region. This data is address level data and must be geo coded, which means it can be difficult to acquire. Once the center receives this data, they must validate the model and check if it works in high-risk areas.

Cook Children's also acquires data from the Department of Family and Protective Services (DFPS), where it receives data on all crimes and offenses. Cook Children's has agreements with the department and goes through a process of data protection in gaining access to this data. There is strong security and the process is approved by the Institutional Review Board (IRB) to ensure proper procedures. Additionally, this data does not contain actual names or information that could pose a privacy concern. Finally, Cook Children's also uses data

purchased from Buxton, a market research and consumer analytics company that provides information about households that fall under the poverty line. This information is determined using a variety of social, demographic, and financial factors relating to the household. The original map that Cook Children's published included poverty at the address-level.

Challenges

One challenge identified by Cook Children's is the issue of getting people to understand the work they are doing, since this type of predictive analytics in this area has not been done before. One concern that people frequently have is that this risk terrain modeling looks at individuals, when in reality the model is finding high-risk areas. A challenge for Cook Children's is convincing people that it is not targeting specific people through the model, but identifying areas where child maltreatment is likely to occur so that action can be taken to prevent occurrences. Cook Children's works at the legislative level with the state to develop policies to prevent child maltreatment from occurring based on the hotspots identified.

Another challenge Cook Children's faces is the difficulty of getting data, especially in a timely manner. The process requires making sure the data is current, sifting through data, finding who owns the data, and effectively communicating how the data will be used in order to convince them to share it. Therefore, Cook Children's most significant challenges involve logistics and the social aspect of communicating the goals of the predictive model and preventing misunderstanding among community stakeholders.

Security

To ensure that the data is protected, organizations, government agencies, and other providers of data only send data to Cook Children's without personal information included. The data contains the fields required to conduct data analysis and analyze the necessary factors without including any personal data. Cook Children's goal in this analysis is not to target specific individuals but highlight certain areas identified as high-risk hotspots so that resources can be appropriately targeted. The center's internal systems department reviewed the security data agreements and confirmed that Cook Children's have the processes required to maintain a proper level of security throughout the stages of data analysis.

Success Metrics

On a high level, Cook Children's can get a sense for the model's potential by looking at the hotspots identified based on risk factors and identify which features are characteristic of hotspots versus non-hotspot areas. One specific metric that provides a measure of the model's success is predictability rate. This can be observed in terms of the model as a whole, as well as per risk factor by looking at each risk factor's ability to predict hotspots accurately. Cook Children's can also measure the success of its predictions by using the previous year's model to predict the following year, and then measuring the accuracy once that year's cases have been identified. For example, the 2014 model was used to predict areas where child maltreatment will occur in 2015, and correctly predicted 98% of the future cases. Figure 2 below illustrates the 2013 predictions of hotspot areas overlaid with the 2014 locations where abuse actually occurred, displaying the accuracy of this modeling technique in predicting locations.

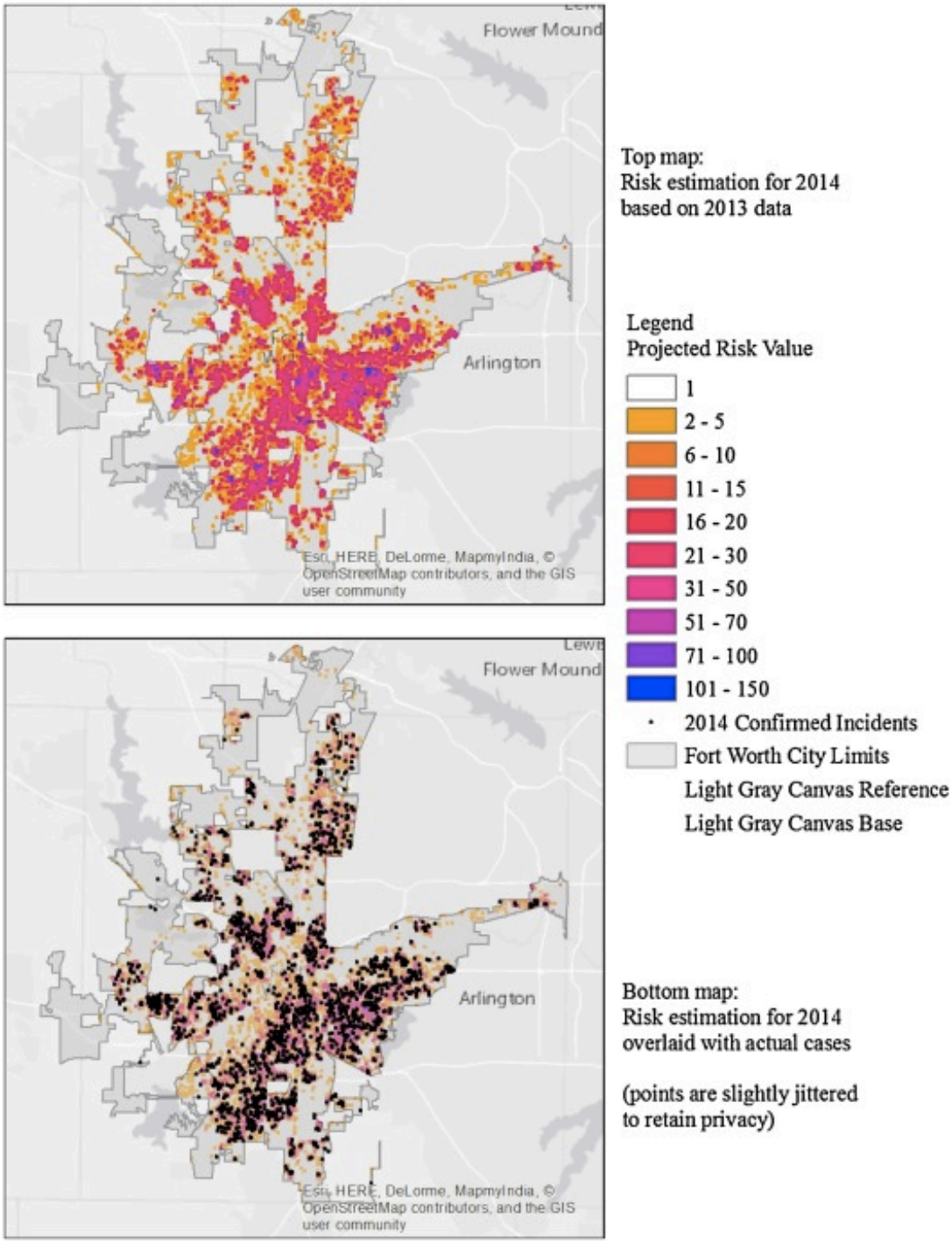

Figure 2. 2013 model predictions overlaid with 2014 actual incidents. The images above illustrate the accuracy of the 2013 RTM model in predicting the actual locations where incidents occurred in 2014 (Child Abuse & Neglect, 2016).

Additionally, the success of Cook Children's initiative is illustrated through the impact of these predictions on the community's resource targeting. Cook Children's goes to different communities and resource organizations that address domestic violence to refocus and target resources based on the areas identified through the model. It can be difficult to track overall community impact due to the number of changing factors, but comparing the metrics before and after it works with organizations to refocus resources can provide Cook Children's with an estimate of the reduction in child abuse.

Next Steps

Going forward, Cook Children's plans to focus on how data can be used to assess individual prevention programs. Currently, surveys are the main source of information used to track progress, but the center hopes to start using data-driven strategies to measure the success of these programs. Additionally, Cook Children's hopes to identify objective metrics that can be used to measure child abuse, citing that a significant challenge is tracking the impact of initiatives due to the number of factors that can influence changes in child abuse occurrences.

Case Study 3: IBM Research – Africa

Organization Overview

IBM is one of the world's largest technology companies, providing a wide variety of computer hardware, software, and other technical services. IBM's products include areas such as cloud computing, analytics, mobile applications, internet of things, security, and IT infrastructure. IBM also offers technology and business consulting services to help design and build systems across various industries. In addition to the products and services it offers, IBM has developed

one of the leading corporate research labs, known as IBM Research, with over 12 labs across the globe. IBM Research is made up of thousands of researchers dedicated to solving the world's most challenging and pressing issues (IBM, 2017).

IBM Research – Africa is the 12th research lab founded by IBM, with locations in Kenya and South Africa. IBM Research – Africa focuses on developing innovative solutions and tapping into new business areas to promote economic growth. The research labs also aim to improve the quality of life across the continent, in areas including water access, healthcare, and financial inclusion. Healthcare is a major focus in Africa, especially after the 2014 Ebola outbreak and its aftermath. During the outbreak, IBM played a significant role in utilizing big data technology to track population movement and contain the spread of the disease in West African nations (IBM Research, 2017).

Data Analysis

To understand the potential of big data in directing public health needs, the 2014 Ebola outbreak in Africa provides an illustrative example. The largest Ebola outbreak to date occurred in 2014 in West Africa. Not only did the epidemic lead to thousands of deaths as the disease spread across the region into heavily populated urban areas at a rapid rate, but it has also continued to put a huge strain on both the economies and healthcare systems of these countries. The use of information technology and tools that facilitated communication and data sharing played a large role in efforts to contain the outbreak. IBM led an effort to address the outbreak, bringing together a variety of companies, research organizations, government agencies, hospitals, and nonprofits to join forces to create innovative solutions based on efficient information sharing. According to IBM, collaboration in data collection and the data-driven solutions that resulted

improved the emergency response and made the country better equipped with the information necessary to contain the outbreak (USAID, 2014).

Mobile health initiatives in particular played a large role. IBM developed a cloud-based big data platform to aggregate data from different sources. Working closely with the Sierra Leone government, Centers for Disease Control and Prevention, and local organizations, IBM Connections platform aggregated SMS data with location and demographic data, with the ultimate goal of creating an open access data storehouse available to all and using it to generate valuable insights. The telecommunications company Airtel created a toll-free number through which citizens can send SMS messages related to Ebola, and then Echo Mobile, a Kenya-based startup, anonymized this data. Combined with GPS data, the SMS data was stored in the platform and analyzed to help public health officials in tracking the disease's spread and assisting the transfer of medicine and resources between clinics. Doctors and health workers benefited from this improved efficiency, as test results could be scanned to and from tablets or uploaded to emergency databases, and text message alerts of test results were sent to field teams on their phones. Mobile health platforms drastically brought down infection reporting time and, as a result, emergency workers managed to reach many of those who came in contact with the virus outbreak.

Through advantages like real-time communication and GPS tracking of movement, technologies that facilitated data gathering and analysis on a large scale allowed for efficiency in treatment and better understanding of broad trends, as IBM claims. By interpreting data in real time, analytics can reveal trends and conditions that may otherwise have been missed, thus allowing for closer monitoring of disease patterns. IBM cited that improved understanding of the outbreak and knowing where to focus resources reduced wasteful spending, and more efficient

coordination of resources, medicine, and health workers eliminated inefficiencies and brought down the number of deaths. Thus, collaborative efforts to gather data, and the resulting insights that were generated through compiling and analyzing large datasets, allowed the country to make use of big data and apply it towards saving lives outbreak.

Data Sources

In order to create a cloud-based big data platform, IBM aggregated data from a variety of sources. As mentioned above, IBM partnered with local governments, Centers for Disease Control and Prevention, and local organizations, hospitals, and nonprofits to collect data for the IBM Connections platform. Combined with SMS, location, and demographic data, this platform provided insight on the spread of Ebola and allowed resources to be allocated to the areas that needed them most. A major aspect of this data collection was the SMS text messages and voice calls that enabled citizens to report issues relating to Ebola. By collecting data from citizens, the government and partner organizations were able to include citizens in the data collection process and understand the issues from the communities that were directly affected, allowing them to keep these concerns in mind while developing strategies for containing the spread of the disease (IBM News Room, 2014).

Challenges

One significant challenge IBM encountered while organizing this initiative was dealing with multiple stakeholders across the data chain. Receiving different information from different sources, for example organizations versus workers on the ground, can make it difficult to gain feedback and assess the situation. In order to understand the process, identify who the users were,

and analyze the context in which these users worked, IBM had to make sure all voices were heard and all sides were taken into account. As IBM made clear, technology of this nature is the most beneficial when it is built based on the user's pain points, so focusing on feedback from all stakeholders was crucial in making improvements to the technology.

Another challenge involved actually deploying the system, as well as ensuring that it was sustainable in the long run. Given that the platform was based on SMS data and relied heavily on the engagement of the Sierra Leone community, dealing with resource constraints was a major concern. The success of the platform relied on the ability to receive SMS messages from citizens, run analytics to categorize messages, and analyze trends from these messages along with other demographic data collected. From this analysis, IBM directed resources to different health service locations based on demand so that those who needed help had access to care when they were sick. However, it was not always possible to rely on people being able to get to these facilities, even if resources were made available for them. At times, issues like these complicated the process. The data analysis may pinpoint the areas with the highest risk of disease, and resources may be allocated based on this analysis, but the solutions may not reach all individuals due to unforeseen complications. While this was an issue throughout the process, IBM claims that the nature of the technology and the close communication it allowed for with and between citizens enabled it to be successful despite these difficulties. Additionally, the community engagement aspect of the platform contributed to its sustainability by facilitating the exchange of information between Sierra Leone citizens and the government.

Security

In order to ensure that the data was kept private and secure, IBM followed international guidelines and corporate policy regarding data security. Additionally, the SMS data that citizens sent which was aggregated in the platform was kept anonymous. Echo Mobile, a startup in Kenya, anonymized the data after Airtel set up a toll-free number for citizens to send SMS messages to. As a result, citizens could send messages without their personal information being stored in the database. This information provided insight on the spread of Ebola and allowed researchers to track the movement of the population and coordinate medical resources without gathering personal data on specific individuals. Additionally, the IBM Connections technology has been used in many other situations and was proven to be a secure and reliable platform. The platform allows for secure document sharing, and only authorized users can access the data stored in the digital cloud (IBM News Room, 2014).

Success Metrics

In terms of measuring the success of the initiative, IBM conducted both pre-deployment and post-deployment evaluation to understand the initial goals of the project and its ultimate impact. In the pre-deployment evaluation, the focus was on understanding what the goals are, establishing a timeline, and planning resources. The post-deployment evaluation analyzed how the new system changed the way the issue was addressed, and compared the final results to the pre-deployment expectations. A major focus of the post-deployment evaluation was to gather feedback on the project to understand how this solution can be applied in new situations. Additionally, the platform involved opening a channel for communication with citizens so that

IBM could gain feedback from the general public on how to provide innovative technology and make improvements to its strategy for fighting Ebola.

Next Steps

Going forward, IBM plans to continue developing this platform and finding different applications for it. Many countries in Africa can benefit from a similar technology, and IBM has started sharing the cloud-based platform with governments and nonprofit organizations. IBM plans to extend these efforts and examine mobile phone signal data with the goal of tracing population movement on a large scale to help researchers predict the spread of disease. Additionally, IBM intends to grow the open data repositories to help mitigate disease epidemics. For example, IBM volunteers set up the Ebola Open Data Repository, a cloud-based repository that organizations across the world contribute to that will provide governments and researchers with accessible data. Open data can be valuable in collaborative efforts that address large scale issues like disease outbreaks, and data repositories like this one will help drive innovation in future projects (IBM News Room, 2014).

Case Study 4: Austin Area Sustainability Indicators

Organization Overview

The Austin Area Sustainability Indicators began in 1999 as an initiative aimed at tracking metrics of sustainability in Austin. It has expanded over the years and now measures sustainability across six counties in the Greater Austin area. The RGK Center for Philanthropy and Community Service at The University of Texas at Austin's LBJ School of Public Affairs, took over the project a year ago with the goal of connecting philanthropy and nonprofits through

the use of data. The RGK Center's mission is to train future nonprofit and philanthropic leaders, and it believes the Austin Area Sustainability Indicators project serves as a data-driven approach to addressing the major issues in the Austin area.

The project's vision is to use evidence-based collective impact to improve the Austin community, promoting collaboration across philanthropic and nonprofit organizations, social entrepreneurs, and citizens. Its goal is to provide an innovative platform to analyze sustainability metrics to make informed decisions by monitoring trends in a variety of areas, from public safety and healthcare to education and the environment. The RGK Center collaborates with the Austin Community Foundation on their Understanding Austin initiative, which aims to understand the region and know where to focus philanthropic resources. The data from Austin Area Sustainability Indicators is used in this initiative to highlight new opportunities in philanthropy and promote collaboration across the community (Austin Area Sustainability Indicators, 2017).

Data Analysis

The Austin Area Sustainability Indicators project defines sustainability broadly, analyzing data across a multitude of areas. These areas include: demographics, public safety, education and children, social equity, engagement, economy, environment, land use and mobility, health, and food. Analysis is conducted in all these areas to draw attention to the pressing issues in Austin, providing insight on where to focus philanthropy efforts. The project initially began as a type of data report, but now focuses more on data visualization and using the data to catalyze action. According to the interviewee, data visualization strategies are essential in conveying information to stakeholders in a way that the findings are easily understood and geared towards the intended audience. After the data is collected and analyzed, reports are generated using

Tableau software, which allows for user-friendly depictions of large amounts of data. These graphs and charts are available through the Austin Area Sustainability Indicators website, where people can read about the major issues and the data analysis conducted. Given the large volume of data analyzed, data visualization transforms this complexity into recognizable patterns and conveys a message in ways that spreadsheets and simple data reports may not.

For example, education and children is one area that the project seeks to provide insight on through data analysis. One factor analyzed within this category is the quality of childcare available to families. Finding a high quality affordable childcare facility can be a challenge for many families. While the Austin area has nearly two thousand childcare facilities, only 6% of these are accredited and only half are licensed. The available capacity of childcare facilities in the area has remained and continues to stay below the population of children ages 0 to 6 years old. The graph below, included on the Austin Area Sustainability Indicators website, provides a snapshot of this discrepancy between child population and facility capacity. Although this is only one example of the type of data the Austin Area Sustainability Indicators project analyzes, it illustrates the information the project provides that can be used by organizations in the community working to address these problems (Education & Children, 2017).

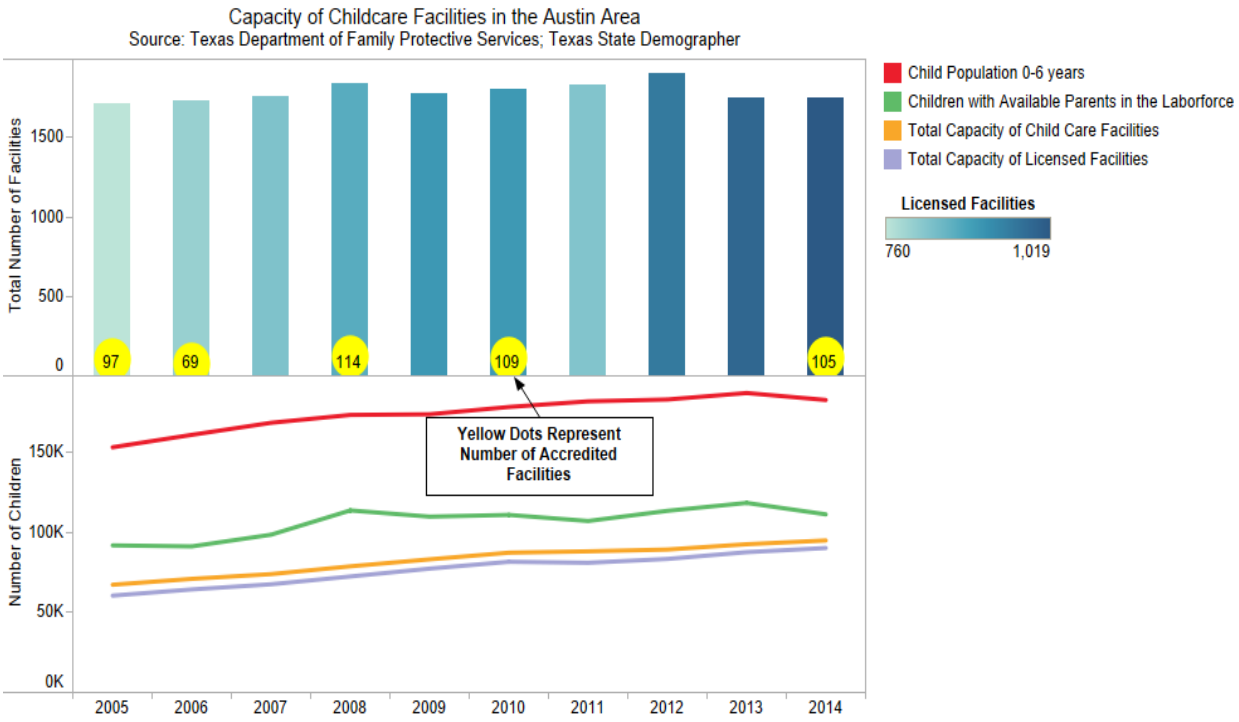

Figure 3. Childcare facilities and child population in Austin. The graph above illustrates the capacity of childcare facilities in light of the child population in Austin, highlighting the need for increased capacity and availability of affordable childcare (Education & Children, 2017).

The Austin Area Sustainability Indicators project receives funding from a variety of donors, including the RGK Center and the Austin Community Foundation. Many of these donors are organizations and foundations in Austin that will benefit from the project's data analysis. This creates a sustainable financing mechanism, since funders provide financial support for the project and receive useful data in return. For example, the Austin Community Foundation is an organization that promotes philanthropy in Central Texas and provides funding and support to mission driven organizations. The Austin Area Sustainability Indicators project's data analysis provides information to them on the major issues in Austin, helping to connect them to nonprofit organizations that address these challenges. By providing funding for Austin Area Sustainability

Indicators, Austin Community Foundation supports the project while also gaining valuable insight from the data analysis its funding makes possible.

Data Sources

The Austin Area Sustainability Indicators gathers a total of 138 data metrics to use for analysis on the variety of topics it explores. To collect this data, the project uses both primary and secondary sources of data. One value-adding piece of the project, according to the interviewee, is the collection of primary survey data. Austin Area Sustainability Indicators surveys local residents in the six county regions about their attitudes and beliefs around the different issues, such as poverty, climate, and affordability, and complies these survey answers into datasets. In addition to this primary data, they collate data from secondary sources including the U.S. Census Bureau, crime statistics, government agencies, and other publicly available sources of data.

Challenges

One of the biggest challenges Austin Area Sustainability Indicators faces is at the data collection stage, specifically in terms of designing the database. The goal is to structure data in a way that they can do meaningful analysis on it with data coming in from various sources and at different levels, such as county versus city data, and at different units of analysis. This functionality requires designing a relational database that speaks to all these dimensions. For example, being able to combine secondary and primary data automatically would significantly improve efficiency but would require the secondary data to be structured in the proper way, which is not always the case. Additionally, the collection process itself could be more

streamlined and efficient, such as having the database automatically download data. As the interviewee cites, having a database with a higher level of automation and efficiency would allow for further analysis, such as predictive analytics.

In addition to data collection challenges, Austin Area Sustainability Indicators also faces challenges in terms of technology. Currently, Excel and Tableau are the main tools used for data visualization. They believe they could be using this software in more ways, for example taking survey data and developing filterable data visualization, such as a feature that allows for filtering the factors contributing to crime rates in low-income neighborhoods. Additionally, they plan to make use of IBM'S SPSS software for predictive analytics in expanding their data analysis. Thus finding new ways to use existing software as well as utilizing new technologies will be the next step for the project.

Security

The Austin Area Sustainability Indicators project is guided by the Institutional Review Board (IRB) for all its data collection and management. The secondary data it sources is already aggregated and is publicly available, so privacy is mainly a concern for the primary survey data it collects. To maintain privacy, the survey data collected is anonymized and only the respondent identifiers are kept in the data file used for analysis. The data file is password protected and only shared with the researchers who are approved through the IRB approval. Additionally, analysis is never done on specific respondents since all aspects of the data report are aggregate. Survey respondents are selected randomly and are not required to provide personal information like names and addresses, with the exception of zip codes and county of residence, which are often necessary for data analysis purposes. The main risks with survey data are privacy and

confidentiality, and the IRB process ensures that researchers follow proper standards for protecting information.

Success Metrics

Given the collaborative nature of the solutions to the issues the Austin Area Sustainability Indicators projects measures through its data, there are different ways in which success can be measured. There are both direct and indirect changes that occur due to the insight provided by the data analysis, so pinpointing the direct impact can pose a challenge. Generally, researchers can look at fluctuations in data and see the needle move on many of these issues. For example, when measuring poverty rates, they can look at the different data metrics relating to poverty and track where specifically the data has changed in a positive way. After understanding the impact, they can then create feedback to understand the relationship between different factors. Another indicator of success is the strength of relationships with partner organizations. Although it is separate from the data, they believe these relationships are an important factor to consider and shed light on the impact of the project on fostering philanthropy and collective impact throughout the community.

Next Steps

Going forward, the Austin Area Sustainability Indicators project plans to conduct analysis on the sustainability for the entire hill country, which includes seventeen counties compared to the current six counties they focus on. The project's goal is to establish a database or standard protocol so that this analysis can be replicated in different places and on different scales. Additionally, there are a variety of side projects and different directions that have come

out of the data that the project plans to collaborate with. For example, the project has plans to partner with a local radio station on a series relating to civic engagement and social media. By providing data to the community, the project gains new partners and new opportunities for collaboration on projects that it plans to pursue.

Analysis

As the four case studies have illustrated, organizations face numerous challenges when it comes to using data-driven strategies towards social impact. Although these organizations have been successful in leveraging big data to improve the quality of society, they have had to overcome many impediments in order to do so and continue to encounter challenges despite their success. This section will first provide an overview of the major challenges these four organizations dealt with, before analyzing the strategies that have proved effective for overcoming these barriers.

Challenges

One difficulty cited by all four organizations was the issue of acquiring data, specifically negotiating data sharing agreements. All four emphasized the importance of data sharing, as developing partnerships or agreements with entities that provide data is an essential aspect of the data collection process. Children's Optimal Health and Austin Area Sustainability Indicators rely heavily on data from other organizations and government agencies in Austin. IBM's platform required the collaboration of numerous companies and groups in Africa in order to aggregate the various types of data needed for disease tracking. Cook Children's relies on data from other organizations, and a challenge it has dealt with is gaining timely and correct data. Given the

importance of data sharing for these organizations, developing proper sharing agreements is a crucial step in guaranteeing that the variety of data required is collected.

Another challenge the organizations encountered is ensuring that the intended audience understands the meaning of the data. In any data analysis, there are stakeholders across the data chain, from collection to analysis to users. Keeping in mind the end users throughout the process is essential in order for the data analysis to catalyze action. For example, Cook Children's struggled with educating the public about the goals of its predictive analysis and dispelling any misconceptions about the data targeting specific individuals. Both the Austin Area Sustainability Indicators project and Children's Optimal Health place a high importance on data visualization and finding ways to convey data in a user-friendly manner. Another challenge relating to understanding users is the issue of building technology based on the needs of those who interact with it. For example, IBM's Africa initiative dealt with this challenge in designing the cloud-based platform and data collecting mechanism for aggregating the data. Therefore, ensuring that end users are kept in mind throughout the design and analysis stages of the process, despite the complexity of the data, is a common challenge.

Finally, success metrics and scalability are also issues the organizations faced. In terms of finding metrics to measure success, defining precise metrics to evaluate impact is a challenge that organizations committed to social impact encounter. In particular, organizations that use data analytics to provide insight to community stakeholders and encourage action run into this difficulty since the far-reaching impact of these initiatives can be difficult to trace. Children's Optimal Health deals with this challenge as it attempts to measure the impact of its work, given this issue of tracing the direct effects of the maps. Relating to this obstacle is the question of scalability, specifically building systems that are scalable and can be made to expand the impact.

All the organizations focus on sustainability, developing solutions that are long-lasting, allow for growth, and have the potential to be applied in new ways. The Austin Area Sustainability Indicators project strives to improve its technology and develop an advanced database that will enable data analysis and collection on a larger scale in order to impact a larger community. Similarly, Cook Children's and Children's Optimal Health continue to expand the projects they work on and apply their analysis techniques to reach new areas. IBM also develops strategies to make use of its technology in multiple African nations and develop a larger scale database that will assist with disease outbreaks in the long term.

Strategies

In light of these challenges, what strategies can be utilized in order to overcome the barriers and accelerate the use of big data towards social innovation? A number of lessons can be gathered from these case studies to guide efforts in developing a plan for improvement. This section will focus on the strategies these organizations and other successful initiatives have employed in order to achieve large-scale impact through big data, as well as areas for improvement, offering solutions that can be used going forward.

Community Engagement

The use of big data is generally associated with companies, organizations and governments. However, the role of coordinating and facilitating big data strategies is not limited to professionals. In many cases, engaging the community and allowing citizens to assist in efforts to collect and analyze data can be an effective strategy (United Nations, 2014). In order to maximize the benefits to the communities the data is intended to help, efforts to bring about

change need to begin at the community level. The government and large organizations undoubtedly play a key role in moving big data efforts forward, whether it is through advancing technologies, initiating projects, or developing policies, but improvements cannot come without involvement from communities. The role citizens can play is often overlooked despite the fact that they are a valuable tool for providing data, especially when the issues addressed directly impact their community.

All four of the organizations made use of community participation in their data initiatives. Austin Area Sustainability Indicators, for example, engages the community as part of the data collection process by collecting primary data from surveys of people living in Austin. Additionally, the data report is made available to everyone, so even individuals in the community can take part in developing solutions to community issues by understanding the community's needs based on the data. Both Children's Optimal Health and Cook Children's work with individuals and community partners as well, producing data reports and maps that are intended to engage the local communities by highlighting the problem areas and assisting with the development of solutions.

When it comes to advancing the use of big data and increasing its potential in developing countries, involving the community is a necessary step towards more widespread use of data-driven strategies. In Sierra Leone, IBM's SMS platform provided citizens with a means for communicating their issues through texts to public health officials, who then coordinated with government agencies and clinics. By allowing citizens to play an active role in data collection, IBM and the Sierra Leone government created an efficient process for aggregating useful data on a large scale. Governments in other countries seeking to better utilize big data in public health initiatives would benefit from a similar community focused approach.

Another role that citizens can play is in the process of measuring the impact of initiatives. Organizations should consider obtaining feedback directly from the community and engaging citizens in the feedback process. Rather than rely on metrics that come from the other end of the data chain, organizations can go directly to the source and find out from the people who are directly impacted by these initiatives. Given that measuring impact and finding success metrics is a challenging task that all four of the organizations identified, as well as an obstacle that many social sector organizations face, the role of citizens should not be overlooked.

In countries where governments have not initiated data movements, citizens have developed their own networks for data sharing and analysis. Communities in Nepal and Uganda created organized efforts to bring together multiple stakeholders, combining citizens, agencies, and organizations to collect and interpret more data. By uniting the community in improving the amount and quality of data, these initiatives aim to harness the power of data in developing innovative solutions to social problems. The goal of these projects is to accelerate the use of data-driven strategies for the country's development, which they attempt to accomplish through coordinating the exchange of data, raising awareness of the potential for increased data sharing, and facilitating joint efforts to encourage collaboration between different stakeholders (Open Nepal, 2017).

Currently, Open Nepal, the project started in Nepal, has over 200 sets of data that have been compiled, shared, and used in analyzing various issues, spanning from health to education to energy consumption. In terms of public health, for example, these datasets have allowed for gaining insight on topics such as preventing malnutrition, treating mental health, and improving access to contraceptives (Open Nepal, 2017). These initiatives demonstrate the potential for community-driven approaches for accelerating the use of data on a large scale, when countries

have not yet developed widespread efforts. Engaging the community in efforts to make use of data promotes the culture of data sharing needed for fostering collaboration. Citizens playing a role in data-driven solutions, especially when they voluntarily release their own data, helps promote open access and more transparency in data.

Another major advantage of enlisting the help of citizens, specifically in developing countries, is the potential for mobile phones as a source of data. The rapid growth of mobile phone usage in the developing world has allowed for large amounts of information collection and improved understanding of population needs. This concept will be discussed in later sections to provide insight on the importance of different types of technology in improving the reach of data-driven strategies.

Scope and Scalability

As these case studies have demonstrated, a clearly defined project scope is a key factor in the success of an initiative. Regardless of project size, having a precisely defined scope is necessary in order to determine data sources and create actionable insights. Without this clarity from the beginning, coordinating resources and time can create difficulties later on in the process and impede the progress of a project. Developing a successful and sustainable solution requires starting with the right question, a task that all four organizations focused on initially. Whether developing maps to convey the health needs of Austin's youth or aggregating cell phone data from individuals in Africa to track the spread of disease, these projects defined from the beginning a clear mission, both in terms of purpose and area served.

In analyzing the scope of a project and the scale of its impact, one factor to consider is the size and type of the organization. Organizational structure can certainly play a role in the

scale of initiatives, influencing the area served and the overall impact. A large international corporation and a smaller nonprofit organization are likely to differ in terms of budget and the nature of the funding. However, the size and structure of organizations and the type of funding they receive do not necessarily impact the scale of the work they do. There are a number of reasons for this, which will be discussed below. Scalability is an important concept to discuss within the context of big data, given that advances in data technology can play a role in achieving scalable social innovations.

A significant trend in the social sector in recent years has been a focus on scalability of social innovations, as new products, services, or processes are increasingly being produced and replicated on a large scale to reach more people. In this context, scaling can be considered in terms of the number of individuals whose lives are improved as a result of the social innovation. An innovation has fully scaled when its impact matches its level of demand (Nesta, 2014). In reality, there are many situations where an innovation doesn't reach its full potential. In some cases, scale may not be suitable. There are certain innovations that are too specific to a certain context to be scaled, or innovations that are more impactful on a smaller scale. Even if an initiative or innovation does not scale up, its impact can be significant – scale is not the only measure of success, and there is still a widespread need for developing local solutions to problems. In some situations, however, there is significant potential for a solution to be scaled up yet it remains on a small scale due to a number of factors. Making the most of a solution involves understanding the root cause of the issue being addressed and then organizing the necessary resources to achieve growth. A significant factor in achieving a large scale has to do with the nature of the innovation or problem being solved. Issues that can be applied in new contexts and are simple to implement, for example, are more easily scaled. Timing is another element, as

some innovations take off due to the current trends or complimentary solutions that are available at a given time. Additionally, a large factor impacting an innovation's potential to be scaled is the ability to make use of existing resources, infrastructure, or systems. When an innovation is compatible with structures that already exist, implementation becomes easier (Nesta, 2014). .

As these factors demonstrate, the dynamics of scaling explain why size of the organization does not always correlate with the scope of the problem being addressed. The access to resources and funding that larger companies have may not lead to a greater impact due to the factors that allow for scaling up solutions. Smaller organizations can achieve scale through a variety of strategies, such as sustainable funding mechanisms or through developing partnerships with other organizations to gain access to resources. This trend can be observed through the four case studies, as differences in organization size, type, structure, and mission did not necessarily impact the scale of the projects. Despite these differences, all the organizations began at the local level and focused on addressing certain problems for that local community. Since the beginning of these initiatives, all of them have expanded their impact as they have grown. While these are still locally focused solutions, the organizations have developed models that are flexible and can be expanded and applied in new ways, which allows for potential growth. For example, IBM began with building a platform in one country before expanding to other West African nations, applying similar technology but adapting it to fit the needs of each region. Children's Optimal Health and Cook Children's have both developed analysis tools that can be applied in new geographic areas and can be used to measure different trends or take into account new measurement factors, all strategies that allow for the potential for growth. The Austin Area Sustainability Indicators project is also developing strategies for growth to allow it to expand to seventeen counties from the original six, by creating growth opportunities through

database design and functionality. What these organizations have in common is that the project design allows for the potential for future growth and has been invented within the context of the funding and resources available.

However, there is no question that certain types of funding are more restrictive than others and undoubtedly present a challenge that many organizations deal with. Nonprofit organizations that are funded primarily by philanthropy, receiving grants and donations from donors and foundations, may especially run into issues relating to flexibility. Given the restrictive nature of grants, as this type of funding often provides financial assistance for a very specific project or aspect of a project, nonprofits that are funded through grants are required to establish upfront goals for projects, which does not allow for flexibility throughout the project (Gregory & Howard, 2009). While large companies that fund their own social impact projects have the flexibility of making changes to project features and adjusting the budget to allow for these developments, organizations that are funded externally lack this ability.

With that being said, there are strategies that organizations can utilize to develop a sustainable funding model in order to prevent this restrictive funding and receive adequate financial support to achieve the scale of impact they strive for. One strategy for an organization to develop a sustainable financing model is to obtain funding from entities that will benefit from the work of the organization. This creates a positive feedback cycle, where a project receives funding from a donor who receives something in return from the project and is therefore motivated to continue funding and perhaps provide less restrictive provisions. The Austin Area Sustainability Indicators has successfully made use of this strategy, as many of its funders are Austin foundations that benefit greatly from the data provided by the project. This type of

relationships between organization and donor creates a long-lasting and sustainable source of financial support.

Another example of a strategy that allows for sustainable funding is to develop a business model that reduces the major expenses through alternative, low-cost innovations. Ascend Learning, a nonprofit charter school network, provides an illustrative example of how a smaller nonprofit organization can overcome funding and resources limitations through innovative strategies (Lindgardt et al., 2013). Ascend Learning is based in Brooklyn, New York, and its mission is to provide education to low-income students in order to close the achievement gap. While many charter schools are limited in their sustainability because of high expenses and a reliance on funding through philanthropy, Ascend Learning has developed an innovative model that allows it to fund itself sustainably. The cost of staff is one of the major expenses for most schools, and Ascend Learning developed a business model that allowed it to avoid this issue. Rather than a labor-intensive strategy, Ascend Learning uses educational technology provided by Sabis, a global education network, which helps to detect gaps in knowledge for each subject and grade level. By applying it in this context, the technology is used towards assisting teachers in understanding the needs of students each week, before these knowledge gaps expand, which means that a given teacher can produce improved academic outcomes for a larger class of students without slowing down the curriculum. Because of this cost-saving model, schools can operate without significant expenses and more students can be offered services. According to Ascend Learning, the program's results have been dramatic, as student test scores have significantly improved, and thousands of students have a higher chance of attending college as a result of the schools. The success of Ascend Learning illustrates how large-scale solutions can be achieved without significant financial input. Through the use of technology and by maintaining a

focus on scalability starting from the initial stages of development, the nonprofit was able to develop one of the largest scale charter school networks by designing a low cost business model rather than a resource-intensive solution (Lindgardt et al., 2013). Ultimately, scaling up requires creative, well-planned strategies in order to be achieved. While resources and funding certainly provide some organizations with scaling potential, organizations of any size can achieve large-scale impact.

In terms of gaining access to advanced technology and other resources that form the foundation of big data solutions, different organizations face different levels of accessibility. The IBM research lab in Africa benefited from the resources that a large company like IBM provides, and had the technology for the cloud-based platform available from the company and was able to apply it in the context of disease prevention. This is certainly an advantage for large for-profit companies when it comes to achieving large-scale impact, since access to technology is more readily available. While this is an obstacle that many nonprofits and smaller organizations with limited technology face, there are other means by which these organizations can gain the resources necessary for data-driven initiatives. In particular, establishing partnerships offers these organizations a chance to gain support for technology-driven projects that require infrastructure they otherwise would not have access to, allowing them to overcome this limitation. The importance of establishing partnerships will be discussed in further detail in the following section.

Collaborative Data Collection

As the case studies made clear, social impact through big data analysis requires collaborative efforts to facilitate the exchange of data. Data collection and finding sources of data presents a challenge to social sector organizations, but this issue can be solved through different forms of data sharing. As this section will demonstrate, there are a number of challenges that stand in the way of cross sector collaboration, but there are promising strategies that can enable the exchange of data and make data increasingly accessible. One of these strategies is to establish data sharing agreements, as the case studies illustrate the importance of sharing agreements to gain access to the data required for analysis. IBM established agreements with local organizations to facilitate efforts and build a cloud-based data network. The Austin Area Sustainability Indicators project and Cook Children's both obtain data from organizations and public datasets. Children's Optimal Health also sources data from other entities, establishing data sharing agreements with partner organizations to develop a steady source of data. Children's Optimal Health must collaborate across community stakeholders since health issues are more effectively solved through cooperative efforts. These case studies demonstrate the necessity of data sharing for nonprofits and other mission-driven organizations that address widespread issues that affect communities.

In addition to data sharing agreements, developing partnerships or support from larger entities can also allow smaller organizations or projects to gain additional data and resources. For example, the Austin Area Sustainability Indicators project is funded by the University of Texas, providing it with access to the resources of a large public institution. Similarly, the Center for Prevention of Child Maltreatment is part of the Cook Children's Health Care System, which provides funding and support. Through partnerships like these, organizations can gain access

either directly to data or to resources and funding that can then enable them to acquire data from other sources.

Given the complex nature of social problems, organizations seeking to tackle such issues through data-driven strategies require comprehensive datasets from a wide array of areas. While massive amounts of data are generated, only a small fraction of this data is made available for analysis. Additionally, the data required for analyzing social and health issues resides in different sectors, often requiring data from both the private and public sectors. In many cases, the power of big data remains untapped because data that can be used towards social innovation is kept privately owned (Stempeck, 2014). As a result, big data solutions require cross-sector collaboration in pooling data, and social sector organizations must find strategies for gaining access to data across sectors.

To combat this issue of siloed data, strategies for facilitating the exchange of data must be developed in order to aid individuals and organizations conducting big data analysis in the social sphere. In many cases, this could be a matter of offering the right incentives for matching data supply and demand between different sources. As mentioned above, establishing data sharing agreements or developing partnerships are strategies that the four organizations made use of in order to access the data they required. While these have proved successful in these cases, there are instances where organizations may not have access to such agreements or collaborations, or where solutions require data on a larger scale. In addition to these methods, there are other solutions that organizations can make use of, or that policymakers, private sector leaders, and even the general public can encourage, in order to ensure that data is available to those who need it.

One trend that offers potential for social impact organizations to gain private sector data is data philanthropy. A relatively new concept, data philanthropy is the idea that private sector firms share their data to benefit the public. The same way companies donate money as a form of philanthropy, donating data can result in significant social impact. There are a number of areas in which corporate data could aid in big data analysis. For example, the IBM platform for tracking Ebola demonstrates the value of cell phone data. Cell providers offering anonymized data can provide a valuable tool for disease containment or any issue that involves tracking human travel. Many private sector firms establish corporate social responsibility initiatives, but given the potential for data-driven social impact, more action is needed in order for these solutions to take off. Granted, there are hesitations when it comes to sharing data – privacy protection, loss of profit, and concerns over competitive advantage are the most common (Pawelke & Tatevossian, 2013). While these are valid concerns, there are strategies for sharing data that can mitigate risks, such as anonymizing data and eliminating personal information to ensure privacy. Ultimately, information on all aspects of our lives are tracked every time we interact with devices, and companies opening up anonymized data does not present a threat that cannot be mitigated through responsible, well-planned measures. Withholding data prevents analysis and research that can have significant social benefit, a potential that outweighs the costs. Additionally, this form of philanthropy benefits companies themselves, as companies seek to give back in a way that is meaningful and most efficient. Providing data is a form of social responsibility that can have a large impact and, going forward, encouraging companies to share data can provide the access to private sector data needed to advance the application of big data towards public good.

Another area for collaborative effort is building collections of data on a regional, national, and global scale to make data more readily available. A growing trend in big data collection is

international data collection, as organizations around the globe have initiated efforts to consolidate data into large publicly available datasets. Building collections of data on a global scale is a step towards big data having a greater impact when it comes to informing decisions. By making it an international effort and encouraging collaboration between governments, efforts to collect large amounts of data will more successfully capture relevant information to understand complex problems related to a community's needs. Despite the volume of data generated, there are surprising gaps in what is known about certain regions and populations (United Nations, 2014). Public health provides an illustrative example. When it comes to public health policies, planning requires up-to-date estimations of disease prevalence, health care access, and clinic resources, among other health care factors (Bloomberg Philanthropies, 2016). As a result, innovative data solutions require more complete and current sets of data, which global data banks can make possible. Not only does international data collection increase the amount of data available, but interpreting data in a broad context also makes it possible to better understand a specific community or geographic area.

The goal of the IBM Connections platform, used during the Ebola outbreak in Africa, is to aggregate data from multiple countries to create a more complete set of data. In tracking disease spread and human population movement, larger databases allow for more potential in developing an accurate prediction of disease patterns. Governments can work together with providers of health apps such as this one to accumulate data to inform health policy. An example of the potential for international databases in improving policy is the Global Burden of Disease (GPD) project. Global Burden of Disease compiles data from researchers, governments, and organizations from across the world into a database in order to help policymakers understand the health challenges their country faces. The goal of the information collection is to identify the

major causes of death in every country in the world, taking into account factors that allow for predictions of disease in the long run. A major focus is the analysis of health trends over time, as well as understanding how health varies based on demographic factors like region, income, and ethnicity. The research done by GBD focuses on both the prevalence of diseases as well as their relative harm, attempting to close the gap in health data. The data currently captures information about over 300 diseases in 188 different countries, and is available for free so that policymakers and researchers across the globe can use the information in making public health decisions. With an interactive use interface, the data is easily accessed, edited, and analyzed (IHME, 2017).

As an example, researchers used data provided by GBD to measure the prevalence of childhood stroke and more accurately understand its global burden. By providing a current and accurate estimate to demonstrate the urgent need, better research and prevention strategies can by developed based on their findings (IHME, 2016). Another study examined the effects of air pollution on health in 188 countries to determine the amount of air pollution in countries relative to each other and over a two-decade period of time (Brauer et al., 2016). The applications of the large amounts of data collected and made available through the Global Burden of Disease are widespread, allowing for a more accurate and holistic examination of the causes and impact of diseases.

Collecting data on an international level provides a number of advantages. It allows individuals and organizations from around the world to contribute data, thus creating sets of data large enough to generate insights from (United Nations, 2014). Many issues faced by society require a significant amount of data, which cannot always be gathered on a community or even national level. When analyzing data, researchers can draw from information on other countries facing similar issues to better understand the needs of a certain issue. Finally, global initiatives in

data collection and analysis also help promote a culture rooted in collaboration and data sharing. Collaboration across countries to collect data paves the way for more cooperative efforts in analyzing data, developing new technology, and accelerating innovation in public health strategies.

As demonstrated, there are a variety of strategies that can combat the issues of data access. Currently, a significant obstacle standing in the way of big data for social impact is siloed data and restricted access. Collaboration across the private and public sectors to facilitate the exchange of data, donation of data, and aggregation of large public data sets will set the scene for larger scale big data analysis that requires data from various sources.

Data Visualization and Storytelling

Data visualization and the concept of conveying data meaningfully are noteworthy factors that came up in many of the organizations. From Children's Optimal Health developing neighborhood maps to illustrate health needs, to Austin Area Sustainability Indicators publishing data reports to the Austin community, the data analyses provided by these organizations have a visual element that enables them to communicate their findings. The goal of these organizations is to provide actionable insight and in order to bring about that action, data must resonate with the people who view it. If findings aren't understood, the data will not result in change and the impact will be lost. Given how difficult it can be to understand large datasets, conveying a narrative behind the numbers and using creative visuals including graphs, charts, and other images allows organizations working with data to bring their findings to life and catalyze action. Data visualization and data storytelling, therefore, are important considerations that can improve

big data initiatives, both in terms of communicating findings as well as measuring and conveying its impact.

Given the necessity of considering data analysis in light of its target audience, focusing on creating aesthetically pleasing visuals allows organizations to make their data easy to interpret and more likely to have the intended effect. There are many situations where data analysis does not have the impact it could have simply due to the way the findings are conveyed, as certain methods of communicating data may not effectively get the point across. This concept is supported by neuroscience research, as researchers have studied the human brain in the context of reacting to facts and information. These studies have demonstrated that logic is not the only factor influencing decision-making, as people make decisions based on both reason and emotion (Verweji et al., 2015). Conveying data as a story, painting a picture around the numbers through both text and visuals, is therefore more likely to be memorable and convincing. In doing so, individuals and organizations presenting data can captivate their audience and not only provide them with data they can interpret, but also make them feel motivated to act upon the information.

Data visualization is a growing trend in data analytics, as data analysts transform numbers into appealing visuals that break down complexity and make it easy to identify patterns and trends. Data visualization combines design thinking with analytics to provide visually appealing representations of data to the people that need to see it. These images can range from line charts to heat maps to infographics, and can even include interactive functions that allow users to engage with the data. Color is another factor in visualizations, for example using bright colors to accentuate the main point being made or differentiating categories of data by color. Generally speaking, data visualization refers to any techniques that are used to portray data in a graphic or visual context (Duarte, 2014). Particularly with big data, the high volume and variety

of data demands an efficient way of reducing clutter and presenting data in order to identify trends amidst complexity. A broader concept is the idea of data storytelling, the ability to not only process data but then also communicate the data in a way that is engaging and compelling. Focusing on the context of the data and the appropriate graph or visual can help to convey the message to the data's audience and capture their attention (Dykes, 2016).

Storytelling through data involves more than simply creating aesthetically pleasing visuals of data. To effectively communicate the message behind data-driven insights, it is essential to develop visuals along with a narrative that provides context for the data. Through bringing together commentary and visuals, the full story behind the data can be conveyed in a way that is clearly interpreted and resonates strongly with people. Patterns that may stay hidden in spreadsheets can emerge through data visualizations, particularly when dealing with the complexity of large datasets (Dykes, 2016). Children's Optimal Health provides an example of combining narrative with visualization to effectively communicate data findings, as the organization creates and publishes reporting based on its findings. These reports contain in-depth commentary along with maps outlining health needs among Austin's youth, striking a balance between textual and visual analysis. The Austin Area Sustainability Indicators project also places a significant emphasis on data storytelling. The data reports published on the project's websites include a discussion of findings along with Tableau-generated visualizations that bring the numbers to life to shed light on the major obstacles facing the Austin community. In exhibiting their data analyses in this manner, these organizations present their data in a way that not only allows the Austin community to interpret it but also sparks action, encouraging data-driven conversations and community action.

These same strategies can also be applied when considering success metrics and measuring performance, since nonprofit organizations often have to display their success to funders and donors. Much like conveying the findings of data, visualization and communication strategies can be used to convey the impact of data as well. Although all four of the organizations have found ways to convey their work, many of them cited the issue of finding metrics to measure the impact of their projects as a challenge. Developing a comprehensive system for measuring impact presents a challenge to social sector organizations. Compared to the private sector, where metrics are concentrated on profit, metrics in the social sector are generally less clear-cut due to the variety of impacted groups these organizations report to, including donors, governments, policymakers, and communities. Not only must organizations demonstrate the work they do and at what financial cost, but also the work's impact from different perspectives. Considering the complexity involved in measuring far-reaching impact, finding metrics that fully capture this can prove challenging and complicates the process of evaluating performance (Gregory & Howard, 2009).

Given the nature of funding, it is essential that nonprofits in particular, or any organization that is funded largely by grants and donors, paint a full picture of their work through success metrics. As mentioned in previous sections, community engagement offers one strategy for measuring success, as gathering feedback from citizens can provide insight on the effects of a data initiative. While the metrics themselves will vary by organization and project, data visualization offers a technique to convey the success of initiatives, the same way that it is used to communicate findings. Graphs, dashboards, charts, and other visual depictions of data can be used to demonstrate the impact of data-related projects to funders and to the public in general.

In summary, presentation of data is an important consideration in big data initiatives, as data must not only be analyzed but also communicated effectively. Presenting numbers in a convincing manner allows organizations to send a message through their findings and convey a full story through data. Through storytelling and conveying the full context behind the meaning of statistics and numbers, data analysis can go beyond simply providing information to serving as a catalyst for community action, inspiring people to ask questions and encouraging individuals and communities to act on these findings.

Technology Availability and Access

The choice of technology is another consideration that plays a role in the impact of data-driven initiatives. In determining technology needs, both functional and nonfunctional requirements must be taken into account. As these case studies have made clear, organizations should utilize technology that not only meets the technical needs for analysis but also makes sense within the context of the situation and the users involved. Factors like availability, cost, user needs, as well as future maintainability and scalability should be given sufficient consideration to ensure that the solution is sustainable. There are two issues to consider when it comes to selecting the proper technology: the type of technology used by the organization for analysis, and the technology available to the end users of the data. In making these choices, it is important to keep in mind all the stakeholders across the data value chain to ensure that the technology, including data analysis software, data visualization tools, and types of devices, is the most available, usable, and cost-effective. The case studies demonstrated the importance of making use of technology that is both efficient and available for use, from the perspective of the organization as well as the users. IBM made use of their existing technology to create a cloud-

based platform to aggregate the data collected, and then used mobile phones to collect and track data from citizens, making the most of limited technology available in developing countries. Cook Children's uses Risk Terrain Modeling technology provided by Rutgers University, applying an analysis tool that has previously been used in different contexts towards child maltreatment for the first time and allowing them to analyzing data efficiently while saving costs. The Austin Area Sustainability Indicators project makes use of Tableau for data analysis and data visualization, providing researchers with a valuable tool for analysis while also benefiting the user through the data visualization functionality.

The IBM platform for tracking Ebola provides an illustrative example of how considering technology from the perspective of both the users and developers and making use of technology that is available can lead to successful and sustainable solutions. As IBM developed a big data platform for tracking the spread of Ebola in order to help contain the disease, it faced the issue of building a solution that will be sustainable within the context of developing countries and the resulting limitations in technology and resources. However, by making use of mobile phones, IBM was able to combat this issue by developing a data collection and analysis process that took into account these limitations. In developing countries, a major impediment to the widespread use of big data is the lack of access to advanced technology (United Nations, 2014). As the use of mobile phones spreads rapidly in developing countries, and as certain carriers are more willing to release data, there is significant potential for mobile phones as a solution for data collection in regions where more complex technological developments and infrastructure would be difficult to establish (Montjoye et al., 2014). As the earlier section illustrated, citizens in communities that health-related big data is intended to benefit can play a role in the data collection and analysis process. One of the most effective tools for gathering population data is

through mobile phones, which have become widely adopted by people across the globe. Large amounts of relevant data can be accumulated on a large scale and at a low cost, as the IBM Africa example demonstrates. Given the widespread use of mobile phones in developing countries, mobile data is easy to collect and contains valuable information of various forms

Population data is limited in many areas, but data from the millions of mobile phones are beginning to help fill this gap. Over 70% of the population in middle- and low-income populations own or have access to mobile phones. Among poor populations, the availability of phones is lower but not significantly. Even for people with the lowest income, mobile phones have started to be viewed as a necessity rather than a luxury item, as "spending on mobile service varies little with income, showing an inelasticity that is characteristic of essential goods (Cartesian, 2014)." Thus the adoption of mobile phones, whether it be owning a phone or having access to other people's, is surprisingly high among poor populations. As an example, data from 2012 suggests that out of people in Kenya making less than one dollar per day, 58% of them had mobile access, a number that continues to grow. The rates are similar among developing countries as a whole, as "mean monthly mobile spending only varies minimally across low and lower income groups (Cartesian, 2014)." In low-income countries, most of these phones are used for basic functions such as calls and text messages, but lower costs and improved network speed has resulted in an increased spread of smart phones. There are a number of reasons why those who normally could not afford luxury items have still adopted mobile phones, including improvements to both professional and personal lives (Cartesian, 2014).

With this growth in data available as a result of mobile data collection comes an improvement in our understanding of population needs through data analysis. Mobile phones not only capture a large volume of data but also a variety of data. From GPS signals to location data

to text messages, the data from mobile phones can be analyzed to generate insights on the spread of disease when compiled together, as the IBM example demonstrates (Montjoye et al., 2014). When it comes to epidemiology, location movement is the key to tracking and treating disease. Areas where particular diseases are highly concentrated are well understood; the data that is missing is the human travel patterns – the number of people arriving in certain areas, where they travel from, the frequency of movement. Collecting this data has proved challenging, as public health officials rely on inefficient methods of tracking travel, in some cases even standing outside train stations and other major hubs to physically count individuals. The five million mobile phones in developing countries, however, have the potential to change this. Even though most are inexpensive phones with little capability besides calls and texts, phone activity can be traced to cell phone towers, providing an estimate of location (Talbot, 2013).

Organizations in addition to IBM have been successful through the use of mobile data for big data analysis in developing countries. For example, epidemiologists studying malaria in Kenya have been able to successfully make use of mobile data in treating and preventing the disease. Researchers discovered a specific cell phone tower where people who were making calls were three times more likely than the average to visit Lake Victoria, an area highly concentrated with malaria infections. This information allowed for a new set of predictive models to gain insight on effective strategies for monitoring outbreaks. The data revealed, for example, the most important areas to focus on mosquito control based on population migration patterns. Using data mining tools to develop an accurate picture of diseases through cell phone data avoids the obstacles created from limited technology (Talbot, 2013). As a result, organizations and governments leading big data initiatives towards social impact will benefit from working around technological constraints by making use of mobile phones.

As the IBM and mobile phone example demonstrates, the type of technology chosen to facilitate collection and analysis plays a significant role in the long-term sustainability of big data initiatives. Certain technologies that may not be fitting in one situation could offer compelling advantages in a different context. The nature of data-driven solutions is entirely dependent on the situation, the issue addressed, and the stakeholders involved from the developers to the end users. Technology and software must not only meet the technical requirements for data analysis but also the nonfunctional aspects, including maintainability and scalability, in order to ensure that it is sustainable and replicable in the long term.

Discussion

The purpose of this comparative case study analysis was to examine the primary challenges standing in the way of using big data towards social impact, in order to determine strategies that can be used to overcome these impediments and bring about more widespread use of data-driven solutions. The organizations included in the analysis varied in terms of size, type, funding, and area focus to reflect the diversity that exists among organizations in this space and to take into account the dynamic nature of trends in big data and the social sector. Ranging from a small nonprofit to a large private sector company, these case studies demonstrate that the role of big data is not limited to a specific type of entity, but rather that any individual, group, or organization can achieve substantial social impact through careful planning and strategic use of resources.

The results of the case studies reveal that the major obstacles faced by organizations attempting to make use of big data revolve around collaboration and facilitating the exchange of data and information. Data sharing is one of the most striking issues, as the data necessary for

analyzing social problems is collected and owned by various sources, requiring coordinated measures to ensure that data ends up in the hands of those who need it. Communicating findings in a way that data is properly interpreted also poses a challenge, given the complexity of large-scale data analysis and the potential for data-driven findings to catalyze community action. Additionally, building technology based on the needs of users in order to ensure that solutions are both scalable and sustainable in the long run is another factor that can complicate big data initiatives. In terms of sustainability, a noteworthy obstacle is the difficulty of finding proper metrics to evaluate the success of initiatives in order to communicate the impact in a way that is measurable. When considered together, these challenges provide a depiction of the impediments standing in the way of solutions rooted in big data to become more widespread in the social sector, and the results explain the gap between the current state and the potential for data-driven social impact.

In light of these challenges, this paper proposes a set of strategies organizations interested in using big data towards social innovation can utilize in order to address these obstacles, highlighting the next steps required to advance the use of big data in the social sector. As organizations move forward, they should consider the following recommendations, summarized below, when designing initiatives aimed at applying big data towards social impact:

1) **Community Engagement:** By including citizens in the data collection and analysis process, organizations will benefit from efforts that begin at the community level and are tailored towards the needs of citizens, who provide both data and insight on the issues directly influencing the community.

2) **Scope and Scalability:** In considering the future scalability of an innovation or initiative, organizations should begin with a clearly defined scope and focus on developing innovative models that will allow them to work within the context of their resource constraints.
3) **Collaborative Data Collection:** Data-driven social impact requires a collaborative model given the dynamic nature of social problems and the complexity of their solutions. Through data-sharing agreements, partnerships, and other methods of collaboration, organizations can facilitate the exchange of data to gain access to the data they need.
4) **Data Visualization and Storytelling:** Given how difficult it can be to understand large datasets, conveying a narrative behind the numbers and using creative visuals allows organizations working with data to bring their findings to life and catalyze action.
5) **Technology Availability and Access:** In determining technology needs, organizations should utilize technology that not only meets the technical needs for analysis but also makes sense within the context of the situation and the users involved.

In evaluating the results, certain findings were expected, such as the need for data sharing and coordinated data exchange. Surprising to note, however, was the difficulty the organizations faced in terms of success metrics to evaluate their work. Given the nature of nonprofit funding, finding ways to convey the impact of initiatives is crucial for organizations seeking grants and other types of support. These case studies have made clear the difficulty of identifying specific metrics that provide an accurate portrayal of the impact of big data projects, presenting a significant challenge. Another surprising finding was the similarities across the four initiatives given the stark differences in the organizations. Despite the differences in the size, structure, type

of funding, and issue addressed by each of the organizations, there were significant similarities across the big data initiatives of these organizations. Ultimately, all the organizations cited data sharing and collaboration as a significant feature and challenge they face, as well as the importance of strategies to convey findings and bring about community action. The major takeaway from these shared features is that it demonstrates how this work is not restricted to large corporations and governments, but that even small-scale entities can have large-scale impact through big data.

In evaluating the potential for big data in driving social impact, there are certain risks that must be given proper consideration. The most common concern is the issue of security, particularly in regards to opening up data and making it increasingly accessible. While there is no doubt that security is a serious concern, there are measures that can be taken to mitigate the risk of privacy breaches and security threats. As the case studies exemplified, actions can be taken to anonymize data to ensure that information is shared without posing a threat to personal information, as well as to guarantee safety through the use of secure servers. Security is a concern with datasets and data analysis of any kind, and thus applying big data in the context of social issues compared to other areas does not pose an added risk. Additionally, it could be argued that the benefits to society from data-driven social innovation far outweigh the security risks given that the data is used towards improving and saving lives. With that being said, security and privacy are both factors that should be considered when developing and implementing solutions rooted in big data, and can pose as potential limitations if not properly accounted for.

In addition to security, data ethics is another factor that can limit the validity of projects rooted in big data, given the importance of ensuring that the data collection and analysis process

is ethical and accurate. Analyzing data requires human decisions throughout the process and may not be immune from all subjectivity, potentially providing false illustrations of data. Additionally, there is the worry that data analysis will encourage certain biases and may inaccurately portray results that lead to wrongful actions. While there is no doubt that data-driven approaches have the potential to address a wide range of issues in new ways and offer valuable insight, it is important to consider the ways in which data can be slightly changed to prove a point. As a result, there are many ethical challenges involved in data collection and analysis.

Going forward, this study provides direction in guiding efforts to expand the use of big data in the social sector, offering suggested strategies to develop sustainable solutions amidst the common obstacles impeding such efforts. As technology continues to advance and the world becomes increasingly interconnected, making use of this connectivity and the resulting data that is generated offers potential for using this information towards social change. By incorporating data-driven analysis into efforts aimed at driving social impact, organizations across sectors can facilitate the development of innovative solutions to the problems society needs answers to.

Conclusion

There is no doubt that big data and the massive increase in information offer new possibilities for more informed decision-making in a variety of areas. As demonstrated, the potential for big data in developing innovative solutions to the most pressing issues confronting society cannot be ignored. Data-driven approaches for evaluating the needs of a population offer a clear path to sustainable development, as individuals and communities faced with challenges can greatly benefit from more directed efforts. Given the difficulties that arise when it comes to identifying, measuring, and implementing solutions to the social problems and health needs of

certain groups, making better use of data analytics in understanding the factors influencing these issues opens up new opportunities for greater impact.

As the case studies illustrate, strategies that are effectively coordinated can prove tremendously successful. Much of the data that is needed to generate useful insights is already available. However, gaining access to this data provides a challenge, an impediment that governments, companies, or any organization hoping to adopt big data into their social impact efforts must find ways to overcome. As this paper has suggested, the key issue standing in the way of making the use of big data in such initiatives more widespread is the difficulty of data sharing and the need for facilitated collaboration. Promoting community-led efforts to bring together data, making use of technology that allows for sustainable solutions, and establishing regional and international collaborative datasets are strategies that will make data more available for analysis and advance the effectiveness of big data initiatives. More research should certainly be done in this area, as data-driven social impact remains a new concept and is still in the early stages. As further research is conducted and as an increasing number of social sector initiatives involving big data are developed, more comparative analyses should be used to evaluate case studies. In particular, researchers should consider the strategies for collaboration and tactics for facilitating more coordinated exchange of data in light of new developments and a growing pool of big data initiatives. Although the power of big data towards social innovation remains largely untapped, the information available in today's increasingly digital world offers hope for the future of big data in addressing the needs of society. With the right strategies, data-driven social impact has the potential to improve the lives of people across the globe.

References

A World That Counts: Mobilizing the Data Revolution for Sustainable Development. (2014). *United Nations Data Revolution Group*. Retrieved from http://www.unglobalpulse.org/IEAG-Data-Revolution-Report-A-World-That-Counts

About | Children Optimal Health. Retrieved from http://www.cohtx.org/about/

About GBD. Retrieved from http://www.healthdata.org/gbd/about

About Us | Open Nepal. Retrieved from http://opennepal.net/about-opennepal

Austin Indicators Project | Promoting sustainability through data-based policy solutions. Retrieved from http://www.austinindicators.org/

Big Data, Big Impact: New Possibilities for International Development. (2012). *World Economic Forum*. Retrieved from http://wef.ch/1mvr6P8

Big Data for Development: A Primer. (2013). *United Nations Global Pulse*. Retrieved from http://www.unglobalpulse.org/bigdataprimer

Brauer, M., Freedman, G., Frostad, J., van Donkelaar, A., Martin, R. V., Dentener, F., … Cohen, A. (2016). Ambient Air Pollution Exposure Estimation for the Global Burden of Disease 2013. *Environmental Science & Technology*, *50*(1), 79–88. https://doi.org/10.1021/acs.est.5b03709

Callanan, L., Gardner, N., Mendonca, L., & Scott, D. What social-sector leaders need to succeed | McKinsey & Company. Retrieved from http://www.mckinsey.com/industries/social-sector/our-insights/what-social-sector-leaders-need-to-succeed

Cartesian, & Bill & Melinda Gates Foundation. (2014). Mobile Data for Development. Retrieved from http://blog.cartesian.com/mobile-data-for-development

Children's Optimal Health. (2011). Child Obesity By Neighbordhood and Middle School.

Clinic, C. Cleveland Clinic Unveils Top 10 Medical Innovations for 2012. Retrieved from http://www.prnewswire.com/news-releases/cleveland-clinic-unveils-top-10-medical-innovations-for-2012-131221809.html

COH Story | Children Optimal Health. Retrieved from http://www.cohtx.org/coh-story/

Daley, D., Bachmann, M., Bachmann, B. A., Pedigo, C., Bui, M.-T., & Coffman, J. (2016). Risk terrain modeling predicts child maltreatment. *Child Abuse & Neglect*, *62*, 29–38. https://doi.org/10.1016/j.chiabu.2016.09.014

Data, data everywhere. (2010, February 25). *The Economist*. Retrieved from http://www.economist.com/node/15557443

Data for Health. Retrieved from https://www.bloomberg.org/program/public-health/data-health/

Data Philanthropy: Where Are We Now? | United Nations Global Pulse. Retrieved from http://www.unglobalpulse.org/data-philanthropy-where-are-we-now

Data-Driven Development – Pathways for Progress. (2015). *World Economic Forum*. Retrieved from http://wef.ch/1AGVmGh

Data ethics is a challenge that major foundations can't afford to ignore. (n.d.). Retrieved from http://www.fordfoundation.org/ideas/equals-change-blog/posts/data-ethics-is-a-challenge-that-major-foundations-can-t-afford-to-ignore/

Duarte, N. (2014, April 16). The Quick and Dirty on Data Visualization. Retrieved from https://hbr.org/2014/04/the-quick-and-dirty-on-data-visualization

Dykes, B. Data Storytelling: The Essential Data Science Skill Everyone Needs. Retrieved from http://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/

Education & Children | Austin Indicators Project. Retrieved from http://www.austinindicators.org/education-children/

Gregory, Ann Goggins, & Howard, Don. (2009, Fall). The Nonprofit Starvation Cycle (SSIR). *Stanford Social Innovation Review*.

IBM - United States. (2015, October 1). Retrieved from https://www.ibm.com/us-en/

IBM launches initiatives to help tackle Ebola outbreak. (2014, October 27). [CTB10]. Retrieved from https://www-03.ibm.com/press/us/en/pressrelease/45214.wss

IBM Research | Africa - Locations. (REPLACE). Retrieved from http://www.research.ibm.com/labs/africa/

Lindgardt, Z., Woods, W., Hendren, C., & Thickett, B. (2013, July 18). Business Model Innovation: Ten Lessons from Nonprofits. Retrieved from https://www.bcgperspectives.com/content/articles/business_unit_strategy_innovation_business_model_innovation_ten_lessons_nonprofits/

Madeleine Gabriel. (2014). Making it Big: Strategies for Scaling Social Innovations. *Nesta*.

Markoff, J. (2012, June 25). In a Big Network of Computers, Evidence of Machine Learning. *The New York Times*. Retrieved from http://www.nytimes.com/2012/06/26/technology/in-a-big-network-of-computers-evidence-of-machine-learning.html

Matthew Brack, & Tito Castillo. (2015). Data Sharing for Public Health: Key Lessons from Other Sectors. *Centre on Global Health Security*, 9.

McAfee, A., & Brynjolfsson, E. (2012, October 1). Big Data: The Management Revolution. Retrieved from https://hbr.org/2012/10/big-data-the-management-revolution

McKeever, B. (2016, June 4). The Nonprofit Sector in Brief 2015: Public Charities, Giving, and Volunteering. Retrieved from http://www.urban.org/research/publication/nonprofit-sector-brief-2015-public-charities-giving-and-volunteering

Peter Groves, Basel Kayyali, David Knott, & Steve Van Kuiken. (2013). The Big Data Revolution in Healthcare. *McKinsey & Company Center for US Health System Reform Business Technology Office*, 2.

Roski, J., Bo-Linn, G. W., & Andrews, T. A. (2014). Creating value in health care through big data: opportunities and policy implications. *Health Affairs (Project Hope)*, *33*(7), 1115–1122. https://doi.org/10.1377/hlthaff.2014.0147

Shoemaker, Paul. (2013, February 3). Reconstructing Philanthropy from the Outside In (SSIR). *Stanford Social Innovation Review*.

Smith, J. N. (2015, March). Mystery Killers. *The Atlantic*. Retrieved from http://www.theatlantic.com/magazine/archive/2015/03/mystery-killers/385001/

Stempeck, M. (2014, July 24). Sharing Data Is a Form of Corporate Philanthropy. Retrieved from https://hbr.org/2014/07/sharing-data-is-a-form-of-corporate-philanthropy

Stroke prevalence, mortality and disability-adjusted life years in children and youth aged 0–19 years: Data from the global and regional burden of stroke 2013. Retrieved from http://www.healthdata.org/research-article/stroke-prevalence-mortality-and-disability-adjusted-life-years-children-and-youth

Talbot, D. (2013). Cheap Phones Churn Out Big Data. Retrieved from https://www.technologyreview.com/s/513721/big-data-from-cheap-phones/

Tarrant County is No. 1 in child abuse. Retrieved from http://www.checkupnewsroom.com/tarrant-county-is-no-1-in-child-abuse

The Social Innovation Journey. (2015). *Frost & Sullivan*. Retrieved from https://community.hds.com/docs/DOC-1005318

United States Agency for International Development. (2014). Technical Brief: Use of Technology in the Ebola Response in West Africa.

Using data to predict child abuse. Retrieved from http://www.marketplace.org/2016/04/14/world/using-data-predict-child-abuse-hot-spots

van Panhuis, W. G., Paul, P., Emerson, C., Grefenstette, J., Wilder, R., Herbst, A. J., … Burke, D. S. (2014). A systematic review of barriers to data sharing in public health. *BMC Public Health*, *14*, 1144. https://doi.org/10.1186/1471-2458-14-1144

Verweij, M., Senior, T. J., Domínguez D., J. F., & Turner, R. (2015). Emotion, rationality, and decision-making: how to link affective and social neuroscience with social theory. *Frontiers in Neuroscience*, *9*. https://doi.org/10.3389/fnins.2015.00332

Yves-Alexandre de Montjoye, Jake Kendall, & Cameron F. Kerry. (2014). Enabling Humanitarian Use of Mobile Phone Data. *Center for Technology Innovation at Brookings*.

ZestFinance issues small, high-rate loans, uses big data to weed out deadbeats. Retrieved from https://www.washingtonpost.com/business/zestfinance-issues-small-high-rate-loans-uses-big-data-to-weed-out-deadbeats/2014/10/10/e34986b6-4d71-11e4-aa5e-7153e466a02d_story.html

Biography

Sonali Arora was born in Austin, Texas on July 24th, 1995. She enrolled at The University of Texas at Austin in 2013, majoring in the Business Honors Program, Management Information Systems, and Plan II Honors program. During her time at UT, she was involved in the Undergraduate Research Journal as an editor, worked as an academic tutor, and studied abroad in Buenos Aires, Argentina studying management and emerging markets. Upon graduating from The University of Texas in May 2017, Sonali will work as a business analyst in Seattle, Washington.